Fast estimation of aperture-mass statistics - II. detectability of higher order statistics in current and future surveys

Article (Published Version)

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **508,** 3474–3494 (2021) https://doi.org/10.1093/mnras/stab2819
Advance Access publication 2021 October 1

Fast estimation of aperture-mass statistics – II. Detectability of higher order statistics in current and future surveys

Lucas Porth [1,2]★ and Robert E. Smith[1]

[1]*Astronomy Centre, Department of Physics & Astronomy, University of Sussex, Brighton BN1 9RH, UK*
[2]*Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany*

Accepted 2021 September 24. Received 2021 September 15; in original form 2021 June 10

ABSTRACT

We explore an alternative method to the usual shear correlation function approach for the estimation of aperture mass statistics in weak-lensing survey data. Our approach builds on the direct estimator method. In this paper, we extend our analysis to statistics of arbitrary order and to the multiscale aperture mass statistics. We show that there always exists a linear order algorithm to retrieve any of these generalized aperture mass statistics from shape catalogues when the direct estimator approach is adopted. We validate our approach through application to a large number of Gaussian mock-lensing surveys where the true answer is known and we do this up to 10th-order statistics. We then apply our estimators to an ensemble of real-world mock catalogues obtained from *N*-body simulations – the SLICS mocks, and show that one can expect to retrieve detections of higher order clustering up to fourth order in a KiDS-1000 like survey. We expect that these methods will be of most utility for future wide-field surveys like *Euclid* and the Rubin Telescope.

Key words: gravitational lensing: weak – methods: numerical – large-scale structure of Universe.

1 INTRODUCTION

Weak gravitational lensing by large-scale structure of the light from distant galaxies is a powerful probe for constraining the cosmological parameters and distinguishing between competing models of the Universe (Blandford et al. 1991; Seitz, Schneider & Ehlers 1994; Jain & Seljak 1997; Kaiser 1998; Schneider et al. 1998; Zhang et al. 2007). The first measurements of the correlations in the shapes of distant background galaxies date back more than two decades (Bacon, Refregier & Ellis 2000; Kaiser, Wilson & Luppino 2000; Van Waerbeke et al. 2000; Wittman et al. 2000). Since then, cosmic shear observations have become ever more precise as the coupling of technological advancements and algorithmic developments have enabled us to conduct unprecedented deep optical imaging surveys of the cosmos KiDS,[1] DES[2] and HSC,[3] with current state-of-the-art surveys now mapping thousands of square degrees (Hildebrandt et al. 2017; Aihara et al. 2018; Troxel et al. 2018; Hikage et al. 2019; Asgari et al. 2021). By the end of the decade planned experiments like *Euclid*[4] and the Rubin Telescope[5] (LSST 2009; Laureijs et al. 2011) will map volumes close to the entire physical volume of our observable Universe. In order to make optimal use of these rich data sets, we will need to push forward our understanding and modelling of various physical and measurement effects. In particular, accurate modelling of the non-linear evolution of large-scale structure, including the baryonic physics effects; accurate modelling and correction of the point spread function of the telescope; correcting the bias in the weak-lensing shape estimation algorithms; and accounting for the intrinsic alignments, to name but a few of the main systematics (see Schneider 2006b; Massey et al. 2013; Troxel & Ishak 2015, for a more detailed discussion of these effects).

If the underlying matter density field were a Gaussian random field, then all of the information in a weak-lensing survey would be contained in the shear two-point correlation function. However, physical effects like: the non-linear growth of structure (Bernardeau et al. 2002), the mapping between cosmic shear and galaxy ellipticities (Miralda-Escude 1991), and lensing beyond the Born approximation (Hilbert et al. 2009; Pratten & Lewis 2016; Fabbian, Calabrese & Carbone 2018), all introduce non-Gaussianity in the maps. Furthermore, the non-linear evolution also induces correlations in the convergence power spectrum multipoles, which grow stronger on small scales. This means that the information content of the second-order statistics becomes saturated after a given multipole (Sato et al. 2011; Hilbert et al. 2012; Kayo, Takada & Jain 2013; Marian et al. 2013; Byun et al. 2017). Thus, in order to capture all of the cosmological information available in lensing surveys one must look to the higher order statistics of the shear field (Schneider et al. 1998; Bernardeau et al. 2002; Schneider & Lombardi 2003). Furthermore, owing to the different ways in which the cosmological parameters and nuisance parameters depend on the higher order statistics, the inclusion of such measurements brings with it the further virtue of being able to break parameter degeneracies, e.g. by combining second- and third-order statistics (Kilbinger & Schneider 2005; Semboloni et al. 2011; Fu et al. 2014), or by incorporating the information found in the statistical properties of the peaks in the shear field (Marian et al. 2013; Kacprzak et al. 2016).

★ E-mail: lporth@uni-bonn.de
[1]kids.strw.leidenuniv.nl
[2]www.darkenergysurvey.org
[3]hsc.mtk.nao.ac.jp/ssp/
[4]www.cosmos.esa.int/web/euclid
[5]www.lsst.org

A powerful method to disentangle systematic effects from cosmic shear signals is the E/B decomposition (Crittenden et al. 2001; Schneider, van Waerbeke & Mellier 2002a). At leading order, pure weak-lensing signals are sourced by a scalar lensing potential, which means that their deflection fields are curl free. Equivalently, the ring-averaged cross-component of the shear is expected to be zero (the B mode), while the tangential one contains all the lensing signal (the E mode). Thus B modes enable a robust test for the presence of systematic errors. One method to take advantage of this E/B decomposition is the so-called 'aperture mass statistics' (Kaiser 1995; Schneider 1996; Schneider et al. 1998). 'Aperture mass' ($\mathcal{M}_{\rm ap}$) and 'Map-Cross' ($\mathcal{M}_{\times}$) are obtained by convolving the tangential and cross shear with an isotropic filter function. Therefore, by construction they are E/B-decomposed. Taking the second moment leads to the variance of aperture mass, the third to the skewness, the fourth to the kurtosis, etc.

The standard approach for measuring the aperture mass statistics in data utilizes the fact that, for the flat sky, any n-point moment can be expressed in terms of integrals over the n-point shear correlation functions, modulo a kernel function (Schneider et al. 2002a; Jarvis, Bernstein & Jain 2004). The reason for adopting this strategy stems from the fact that the correlation functions can reliably be estimated in the presence of a non-trivial survey mask. However, for these estimators to be accurate and E/B decomposed, one requires three conditions to be satisfied: (i) the ξ_+/ξ_- correlations need to be measured down to zero separation; (ii) they also need to be measured up to a maximum angular scale, set by the exact form of the aperture mass filter and its angular scale; (iii) the angular bins must be sufficiently fine for the discretization of the integrals to be reliable (Kilbinger & Schneider 2005; Fu et al. 2014). Owing to galaxy image blending, signal-to-noise ratio issues and the finite size of the survey, the lower bound is never possible and the upper bound means that biases can occur due to edge effects. In addition, while the mean estimate is unbiased, the covariance matrix does require one to carefully account for the mask (Schneider et al. 2002b; Friedrich et al. 2016). More recent developments that also make use of the shear correlation functions, while circumventing the issues of E/B leakage on small scales are the ring statistics and COSEBIs (Schneider & Kilbinger 2007; Schneider, Eifler & Krause 2010). While those approaches can, in principle, be extended to higher order statistics, the estimation of the n-point correlation functions turns out to be notoriously time-consuming (Jarvis et al. 2003; Schneider, Kilbinger & Lombardi 2005). Further methods to extract non-Gaussian information from the aperture mass look at its probability density function as a whole (Bernardeau & Valageas 2000; Munshi, Valageas & Barber 2004; Barthelemy, Codis & Bernardeau 2020) or at the distribution of its signal-to-noise ratio peaks (Marian et al. 2012; Heydenreich, Brück & Harnois-Déraps 2020; Martinet et al. 2021).

In Porth et al. (2020), we took a different approach and explored a computationally efficient (accelerated) implementation of the original direct estimator of the aperture mass dispersion (Schneider 1998). Rather than measuring the correlation functions of the shear polar, in this formulation one instead directly measures cumulants of $\mathcal{M}_{\rm ap}$ on a set of apertures and then uses an optimised weighting scheme to average the estimates, along with a restriction on the types of apertures that are acceptable. The present work extends our previous investigation in a number of important ways. First, we construct accelerated direct estimators for the higher order aperture mass moments, including the skewness, kurtosis, etc. Second, we also develop further the multiscale aperture moments (Jarvis et al. 2003; Schneider et al. 2005). These two improvements enable us to better trace the full, harmonic mode, configuration dependence of the convergence polyspectra.

This paper is organized as follows: In Section 2, we introduce key concepts of weak lensing, define the aperture mass and show how its connected cumulants are related to the convergence polyspectra. In Section 3, we revisit the direct estimators for higher order aperture mass measures and construct suitable bases, in which each statistic can be computed in linear time complexity. After investigating the variance of the direct estimators, we give details of our updated algorithm used to perform the measurements. In Section 4, we empirically verify the linear scaling and the measurements of our implementation of the direct estimator on Gaussian mocks. In Section 5, we then apply the estimator to the SLICS simulation suite in order to assess up to which order one can expect to extract information from the aperture mass statistics on a KiDS-1000 like survey. Finally, in Section 6, we summarize our findings, conclude and discuss future work.

2 HIGHER ORDER APERTURE MASS MEASURES FOR COSMIC SHEAR

2.1 Weak gravitational lensing and aperture mass

In this paper, we are mainly concerned with the weak lensing of distant background (source) galaxy shapes by the intervening large-scale structure (for detailed reviews of the topic see Bartelmann & Schneider 2001; Schneider 2006a, b; Dodelson 2003, 2017; Kilbinger 2015; Mandelbaum 2018). The two fundamental quantities describing this mapping from true to observed galaxy images are the convergence κ and the complex shear $\gamma = \gamma_1 + i\gamma_2$, which, assuming a metric theory of gravity, are all derived from an underlying scalar lensing potential. In a galaxy survey, the effective convergence at angular position $\boldsymbol{\theta}$ and radial comoving distance χ can be connected to the density contrast $\delta(\chi\boldsymbol{\theta}, \chi)$ through:

$$\kappa(\boldsymbol{\theta}) = \frac{3}{2}\Omega_{\rm m,0}\left(\frac{H_0}{c}\right)^2 \int_0^{\chi_{\rm H}} {\rm d}\chi' \frac{\chi'}{a(\chi')} g(\chi')\delta(\chi'\boldsymbol{\theta}, \chi') \,, \tag{1}$$

where $\Omega_{\rm m,\,0}$ is the total matter density, H_0 denotes the Hubble constant, a is the scale factor, c is the speed of light, $\chi_{\rm H}$ is the comoving distance to the horizon, and $g(\chi)$ is a weight function related to the normalized redshift distribution ${\rm d}n(z)/{\rm d}z$ of the source galaxies as

$$g(\chi') \equiv \int_{z(\chi')}^{z_{\rm H}} {\rm d}z \, \frac{{\rm d}n(z)}{{\rm d}z} \frac{\left[\chi(z) - \chi'\right]}{\chi(z)} \,. \tag{2}$$

Aperture mass was developed by Schneider (1996) as a technique to estimate projected mass overdensities enclosed within a circular region:

$$\mathcal{M}_{\rm ap}(\boldsymbol{\theta}_0; \vartheta) = \int_{\mathbb{R}^2} {\rm d}^2\boldsymbol{\theta}_1 \kappa(\boldsymbol{\theta}_1) U(|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0|\,; \vartheta) \,, \tag{3}$$

where U is a compensated filter function. In the flat sky limit the (cross) aperture mass can be expressed in terms of a related circularly symmetric filter function $Q(U)$ and the complex shear field γ in its E/B-decomposed basis:

$$\mathcal{M}_{\rm ap/\times}(\boldsymbol{\theta}_0; \vartheta) \equiv \int_{\mathbb{R}^2} {\rm d}^2\boldsymbol{\theta}_1 \gamma_{\rm t/\times}(\boldsymbol{\theta}_1; \boldsymbol{\theta}_0) Q(|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0|; \vartheta) \,, \tag{4}$$

where the tangential and cross-components of the shear field at position $\boldsymbol{\theta}+\boldsymbol{\theta}_0$ with respect to the aperture centre $\boldsymbol{\theta}_0$ are defined as (Bartelmann & Schneider 2001)

$$\gamma_{\rm t}(\boldsymbol{\theta}; \boldsymbol{\theta}_0) + i\gamma_{\times}(\boldsymbol{\theta}; \boldsymbol{\theta}_0) \equiv -\gamma(\boldsymbol{\theta} + \boldsymbol{\theta}_0){\rm e}^{-2i\phi} \,, \tag{5}$$

in which ϕ denotes the polar angle associated with the vector $\boldsymbol{\theta}$. In the absence of systematic errors (B modes) in the lensing data, map-cross should vanish (Schneider et al. 2002a).

For this work, we will make use of the polynomial filter function introduced by Schneider et al. (1998):

$$Q(\theta;\vartheta) = \frac{6}{\pi\vartheta^2}\left(\frac{\theta}{\vartheta}\right)^2\left[1-\left(\frac{\theta}{\vartheta}\right)^2\right]\mathcal{H}(\vartheta-\theta)\,, \tag{6}$$

where ϑ is the characteristic scale of the filter and $\mathcal{H}(x)$ is the Heaviside function, which guarantees that the filter function has compact support.

2.2 A hierarchy of aperture mass measures

One may construct moments of the aperture mass field, and this gives rise to the so called aperture mass statistics. At the two-point level, this gives us the variance $\langle\mathcal{M}_{\rm ap}^2\rangle_{\rm c}(\vartheta_1)$ and at the three-point, the skewness $\langle\mathcal{M}_{\rm ap}^3\rangle_{\rm c}(\vartheta_1)$, etc., where the subscript c stands for the connected cumulant obtained from the moments (Scoccimarro & Frieman 1996). Owing to the fact that the aperture mass is a convolution of the convergence field with a filter function, it is possible to rewrite these moments in terms of their Fourier space counterparts, that is the convergence spectra. For example, for the variance and skewness we have

$$\left\langle\mathcal{M}_{\rm ap}^2\right\rangle_{\rm c}(\vartheta) = \int\frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2}C_{\kappa,2}\left(\vec{\ell}_1\right)\left|\widetilde{U}_\vartheta\left(\vec{\ell}_1\right)\right|^2\,, \tag{7}$$

$$\left\langle\mathcal{M}_{\rm ap}^3\right\rangle_{\rm c}(\vartheta) = \int\frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\cdots\int\frac{{\rm d}^2\boldsymbol{\ell}_3}{(2\pi)^2}(2\pi)^2\delta^D\left(\sum_{i=1}^3\vec{\ell}_i\right) \times C_\kappa\left(\vec{\ell}_1,\cdots,\vec{\ell}_3\right)\widetilde{U}_\vartheta\left(\vec{\ell}_1\right)\widetilde{U}_\vartheta\left(\vec{\ell}_2\right)\widetilde{U}_\vartheta\left(\vec{\ell}_3\right), \tag{8}$$

where $\widetilde{U}_{\vartheta_i}$ denotes the Fourier transform of the aperture mass filter function $U(\theta;\vartheta_i)$ and $C_\kappa(\vec{\ell}_1)$ denotes the convergence power spectrum, and $C_\kappa(\vec{\ell}_1,\vec{\ell}_2,\vec{\ell}_3)$ the convergence bispectrum. These spectra can formally be defined:

$$\left\langle\tilde{\kappa}\left(\vec{\ell}_1\right)\tilde{\kappa}\left(\vec{\ell}_2\right)\right\rangle_{\rm c} = (2\pi)^3\,\delta^D\left(\vec{\ell}_1+\vec{\ell}_2\right)\,C_\kappa\left(\vec{\ell}_1\right), \tag{9}$$

$$\left\langle\tilde{\kappa}\left(\vec{\ell}_1\right)\tilde{\kappa}\left(\vec{\ell}_2\right)\tilde{\kappa}\left(\vec{\ell}_3\right)\right\rangle_{\rm c} = (2\pi)^3\,\delta^D\left(\textstyle\sum_{i=1}^3\vec{\ell}_i\right)\,C_\kappa\left(\vec{\ell}_1,\vec{\ell}_2,\vec{\ell}_3\right). \tag{10}$$

This of course can be generalized to n-point aperture mass moments:

$$\left\langle\mathcal{M}_{\rm ap}^n\right\rangle_{\rm c}(\vartheta) = \int\frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\cdots\int\frac{{\rm d}^2\boldsymbol{\ell}_n}{(2\pi)^2}(2\pi)^2\delta^D\left(\sum_{i=1}^n\vec{\ell}_i\right) \times C_\kappa\left(\vec{\ell}_1,\cdots,\vec{\ell}_n\right)\widetilde{U}_\vartheta\left(\vec{\ell}_1\right)\cdots\widetilde{U}_\vartheta\left(\vec{\ell}_n\right), \tag{11}$$

where the n-point convergence spectrum is defined:

$$\left\langle\tilde{\kappa}\left(\vec{\ell}_1\right)\ldots\tilde{\kappa}\left(\vec{\ell}_n\right)\right\rangle_{\rm c} = (2\pi)^3\,\delta^D\left(\sum_{i=1}^n\vec{\ell}_i\right)\,C_\kappa\left(\vec{\ell}_1,\ldots,\vec{\ell}_n\right). \tag{12}$$

It is worth noting that due to the fact that $\widetilde{U}$ is a sharply peaked filter function in Fourier space, the aperture mass moment on a given scale only carries information about a specific range of wavemodes $\vec{\ell}$ from the underlying polyspectrum. In order to extract more of the information that is available, one needs to compute equation (11) for a large set of aperture radii (Schneider et al. 2005).

2.3 Multiscale aperture mass moments and their correlators

Even if one considers a wide range of aperture radii there will be certain wavemode configurations of the polyspectra that are suppressed when compared with other configurations. This may result in a loss of sensitivity to certain physical effects that are only manifest in the higher order polyspectra, such as those induced by modifications of gravity or primordial non-Gaussianities. In order to combat this, one can further generalize the aperture mass moments in several ways. First, if we choose different scales for the aperture mass filter function, then we get the multiscale aperture mass moments. For the n-point multiscale aperture mass moment, this can be written

$$\left\langle\mathcal{M}_{\rm ap}^n\right\rangle_{\rm c}(\vartheta_1,\cdots,\vartheta_n) = \int\frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\cdots\int\frac{{\rm d}^2\boldsymbol{\ell}_n}{(2\pi)^2}(2\pi)^2\delta^D\left(\sum_{i=1}^n\vec{\ell}_i\right) \times C_\kappa\left(\vec{\ell}_1,\cdots,\vec{\ell}_n\right)\widetilde{U}_{\vartheta_1}\left(\vec{\ell}_1\right)\cdots\widetilde{U}_{\vartheta_n}\left(\vec{\ell}_n\right). \tag{13}$$

Second, if we correlate a set of apertures at different spatial positions in the sky, then one can define the multiscale aperture mass moment correlators (Szapudi & Szalay 1997; Munshi & Coles 2003). There are two special cases where this approach can be applied, the first is the case where the separation of the aperture is directed perpendicular to the line of sight. The second case is where the apertures are placed along the same line of sight, but where different tomographic bins of source galaxies are used to estimate the aperture mass. The former case measures the correlation of the cumulants on the same redshift slice, but at different angular positions. The latter case corresponds to correlating aperture measures in different surveys with overlapping footprints, or between photometric redshift bins within the same survey. As the aperture mass filter carries most of its weight in a compact region surrounding the aperture centre one expects the signal to fall off rapidly for aperture separations that exceed beyond a few times the aperture radius. Generalizing the result of Schneider et al. (1998), we can formally write this as follows:

$$\left\langle\mathcal{M}_{\rm ap}^n\mathcal{M}_{\rm ap}^m\right\rangle_{\rm c}\left(\vartheta_1,\cdots,\vartheta_n,\vartheta_1',\cdots,\vartheta_m';\vec{\Delta}\right) = \int\frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\cdots\frac{{\rm d}^2\boldsymbol{\ell}_{n+m}}{(2\pi)^2}(2\pi)^2\delta^D\left(\sum_{j=1}^{n+m}\vec{\ell}_j\right)\,C_\kappa\left(\vec{\ell}_1,\cdots,\vec{\ell}_{n+m}\right) \times\tilde{U}_{\vartheta_1}\left(\vec{\ell}_1\right)\cdots\tilde{U}_{\vartheta_m'}\left(\vec{\ell}_{n+m}\right)\,e^{i\vec{\Delta}\sum_{j=1}^m\vec{\ell}_{n+j}}, \tag{14}$$

where $\vec{\Delta}$ is a separation vector. Note that for zero separation we recover the $(m+n)$th cumulant. In addition, we can assess the impact of the exponential factor by evaluating the two-point cross-correlation coefficients r_{mn}, which are defined in a similar way to those in Munshi & Valageas (2005):

$$r_{mn}(\Delta) \equiv \frac{\langle X^m X^n\rangle_c(\Delta)}{\langle X^{m+n}\rangle_c}\,, \tag{15}$$

where for our case $X^m = \mathcal{M}_{\rm ap}^m$. In this work, however, we do not consider the cosmological information contained in equation (14), but instead use it to assess how fast the r_{mn} converge to unity – this can be seen as a proxy for how densely apertures need to be sampled within a survey footprint to retrieve all available signal.

3 ESTIMATORS FOR HIGHER ORDER APERTURE MASS STATISTICS

3.1 Direct estimators for the aperture mass moments and their evaluation in linear order time

In this subsection, we concern ourselves with estimators for higher order aperture mass statistics that mimic the original theoretical expressions equation (11) more closely. At first, let us investigate the special case of all the radii being equal.

Consider an aperture of angular radius ϑ, centred on the position $\boldsymbol{\theta}_0$. The aperture contains N galaxies[6] with positions $\boldsymbol{\theta}_i$, complex ellipticities e_i and weights w_i. Then, for a single aperture, one can write down an estimator for the nth-order aperture mass statistic equation (11) as (Schneider et al. 1998; Munshi & Coles 2003)

$$\widehat{M_{\rm ap}^n} = \left(\pi\vartheta^2\right)^n \frac{\sum_{(i_1,\dots,i_n)^N} w_{i_1} Q_{i_1} e_{t,i_1} \cdots w_{i_n} Q_{i_n} e_{t,i_n}}{\sum_{(i_1,\dots,i_n)^N} w_{i_1} \cdots w_{i_n}}, \tag{16}$$

where we defined the shorthand notation

$$\sum_{(i_1,\dots,i_n)^N} \equiv \sum_{i_1=1}^{N} \sum_{i_2\neq i_1}^{N} \cdots \sum_{i_n\neq i_{n-1}\neq\cdots\neq i_1}^{N} . \tag{17}$$

In certain cases, we might use further abbreviations, meaning that $(i_1,\dots, i_n)^N \equiv (i_1,\dots, i_n) \equiv \neq$. On applying the above estimator to the case of $n = 2$, one can easily show that that this estimator is unbiased after averaging over the intrinsic ellipticity distribution, the galaxy positions within the aperture, and finally over cosmological ensembles (Schneider et al. 1998; Porth et al. 2020).

If we were to apply the above estimator given by equation (16) to determine the hierarchy of aperture mass moments, then this naive implementation would appear to result in an estimator that requires of the order N^n operations to compute. However, following our earlier work (Porth et al. 2020), one can complete the sums to transform the estimators into sums and products of linear order terms. In Appendix A, we explicitly show, using elementary means, how one can compute the skewness ($\widehat{M_{\rm ap}^3}$) and kurtosis ($\widehat{M_{\rm ap}^4}$) using linear sums. The results for second, third and fourth orders are as follows:

$$\widehat{M_{\rm ap}} = M_{\rm s,1} , \tag{18}$$

$$\widehat{M_{\rm ap}^2} = \frac{M_{\rm s,1}^2 - M_{\rm s,2}}{1 - S_2} , \tag{19}$$

$$\widehat{M_{\rm ap}^3} = \frac{M_{\rm s,1}^3 - 3M_{\rm s,2}M_{\rm s,1} + 2M_{\rm s,3}}{1 - 3S_2 + 2S_3} , \tag{20}$$

$$\widehat{M_{\rm ap}^4} = \frac{M_{\rm s,1}^4 - 6M_{\rm s,2}M_{\rm s,1}^2 + 3M_{\rm s,2}^2 + 8M_{\rm s,3}M_{\rm s,1} - 6M_{\rm s,4}}{1 - 6S_2 + 3\left(S_2\right)^2 + 8S_3 - 6S_4}, \tag{21}$$

where we have introduced two additional quantities: S_m and $M_{{\rm s},m}$, which are defined as follows:

$$M_{{\rm s},m} \equiv S_m \left(\pi\vartheta^2\right)^m \frac{\sum_{i=1}^N w_i^m Q_i^m e_{t,i}^m}{\sum_{i=1}^N w_i^m}, \tag{22}$$

$$S_m \equiv \frac{\sum_{i=1}^N w_i^m}{\left(\sum_{i=1}^N w_i\right)^m}. \tag{23}$$

[6]Strictly speaking, we select galaxies within the support of the Q filter function of that aperture. For the filter functions, we use in this work the support is always concentric around the aperture centre and linearly scaling with aperture radius. Therefore, we will continue referring to N as the number of galaxies per aperture.

Applying the elementary approach described in Appendix A beyond fourth order rapidly becomes cumbersome, to say the least. We have therefore developed an analytic method for generation of the nth-order estimator decomposed into linear sums. This follows from noting that the sum in equation (16) runs over unequal indices and that one can express any statistic $\widehat{M_{\rm ap}^n}$ as a sum of the power sums equation (22), where the coefficients preceding each term are determined with the help of the complete Bell polynomials B_n. Hence, for the general nth-order estimate one has

$$\widehat{M_{\rm ap}^n} = \frac{B_n\left(-M_{\rm s,1}, -M_{\rm s,2}, -2M_{\rm s,3}, ..., -(n-1)!M_{{\rm s},n}\right)}{B_n\left(-S_1, -S_2, -2S_3, ..., -(n-1)!S_n\right)} . \tag{24}$$

For full details of this derivation, we refer the reader to Appendix B. Here, we only note that each argument that goes into equation (24) is a single sum over the galaxies in the aperture and is therefore independent of the order of the statistic. Using this formalism, we extend our decomposition to fifth and sixth order, giving us

$$\widehat{M_{\rm ap}^5} = \frac{1}{N_5}\left[M_{\rm s,1}^5 - 10M_{\rm s,2}M_{\rm s,1}^3 + 15M_{\rm s,2}^2M_{\rm s,1} + 20M_{\rm s,3}M_{\rm s,1}^2 - 20M_{\rm s,3}M_{\rm s,2} - 30M_{\rm s,4}M_{\rm s,1} + 24M_{\rm s,5}\right] , \tag{25}$$

$$\widehat{M_{\rm ap}^6} = \frac{1}{N_6}\left[M_{\rm s,1}^6 - 15M_{\rm s,2}M_{\rm s,1}^4 + 45M_{\rm s,2}^2M_{\rm s,1}^2 - 15M_{\rm s,2}^3 + 40M_{\rm s,3}^2 - 90M_{\rm s,4}M_{\rm s,1}^2 + 40M_{\rm s,3}M_{\rm s,1}^3 - 120M_{\rm s,3}M_{\rm s,2}M_{\rm s,1} + 90M_{\rm s,4}M_{\rm s,2} + 144M_{\rm s,5}M_{\rm s,1} - 120M_{\rm s,6}\right] , \tag{26}$$

where

$$N_5 = 1 - 10S_2 + 15\left(S_2\right)^2 + 20S_3 - 20S_3S_2 - 30S_4 + 24S_5 , \tag{27}$$

$$N_6 = 1 - 15S_2 + 45\left(S_2\right)^2 - 15\left(S_2\right)^3 + 40S_3 - 120S_3S_2 + 40\left(S_3\right)^2 - 90S_4 + 90S_4S_2 + 144S_5 - 120S_6. \tag{28}$$

3.2 Direct estimators for the multiscale aperture mass moments

In complete analogy, we can write down an unbiased direct estimator for the full multiscale aperture mass moments of equation (13):

$$\widehat{M_{\rm ap}^n}(\vartheta_1, ..., \vartheta_n) = \left(\pi\vartheta_1^2\right)\cdots\left(\pi\vartheta_n^2\right) \times \frac{\sum_{(i_1,\dots,i_n)} w_{i_1} Q_{\vartheta_1,i_1} e_{t,i_1} \cdots w_{i_n} Q_{\vartheta_n,i_n} e_{t,i_n}}{\sum_{(i_1,\dots,i_n)^N} w_{i_1} \cdots w_{i_n}}, \tag{29}$$

where each index runs through all the galaxies within the aperture of the largest radius. In this case, the power sums of equations (22) and (23) do not form a sufficient basis to express these estimators, but we are still able to write down the estimators from elements within the sets

$$X_n \equiv \left\{ X^{(m)}_{(\jmath_1,\dots,\jmath_n)} \,\middle|\, \jmath_i \in \{0,1\}, \sum_{i=1}^n \jmath_i = m \leq n \right\}, \tag{30}$$

where $X \in \{M_{\rm s}, S\}$ and the corresponding elements constitute of multivariate power sums being defined as

$$M^{(m)}_{{\rm s},(\jmath_1,\dots,\jmath_n)} \equiv \left(\prod_{k=1}^n \left(\pi\vartheta_k^2\right)^{\jmath_k}\right) \sum_{i=1}^{N(\jmath)} w_i^m \prod_{j=1}^n \left[e_{t,i} Q_{\vartheta_j,i}\right]^{\jmath_j} ,$$

$$S^{(m)}_{(\jmath_1,\dots,\jmath_n)} \equiv \sum_{i=1}^{N(\jmath)} w_i^m, \tag{31}$$

where $N(\jmath)$ denotes the number of galaxies within the aperture of the smallest radius for which $\jmath_i$ is not zero. Despite the more complicated looking form compared to the equal radius case these estimators can

also be computed in $\mathcal{O}(N)$ time using the $\left|\hat{X}_n\right| = 2^n - 1$ distinct multivariate power sums equation (31) and summing over various partitions P of the set $\{1,..., n\}$:

$$\widehat{M^n_{\rm ap}}(\vartheta_1, ..., \vartheta_n) = \frac{\sum_{m=1}^{n} \sum_{\pi\in P_{n,m}} (-1)^m \prod_{i=1}^{m} (n_i - 1)!\, M^{(n_i)}_{{\rm s},(\vartheta_1(\pi_i),\cdots,\vartheta_n(\pi_i))}}{\sum_{m=1}^{n} \sum_{\pi\in P_{n,m}} (-1)^m \prod_{i=1}^{m} (n_i - 1)!\, S^{(n_i)}_{(\vartheta_1(\pi_i),\cdots,\vartheta_n(\pi_i))}}. \quad (32)$$

In this expression, the combination of the two outer sums run through each partition π that consists of m blocks and the $\alpha(\pi_i)$ denote the value of the α as evaluated from the ith block of the partition. For a motivation of this equation and explicit expressions, we again refer to Appendix B.

3.3 Estimators applied to a large survey

In order to estimate any aperture statistics

$$\mathfrak{M} \in \left\{ \left\langle M^2_{\rm ap} \right\rangle (\vartheta_1),\ \left\langle M^3_{\rm ap} \right\rangle (\vartheta_1), \dots \right\} \quad (33)$$

on a contiguous survey field one can simply place an ensemble of apertures on the field and compute their weighted means

$$\hat{\mathfrak{M}} = \frac{\sum_{\rm ap} w_{\rm ap} \hat{\mathfrak{M}}_{\rm ap}}{\sum_{\rm ap} w_{\rm ap}}\,, \quad (34)$$

where the weights w_i should be chosen to minimize the variance of the estimator. Owing to the linearity of equation (34), if the estimator of a single aperture is unbiased, then so is equation (34). Thus, including more apertures will increase the signal-to-noise ratio of the ensemble estimator.

3.4 Variance of the direct estimators

In order to understand how to weight the apertures, we need to obtain expressions for the variance of the moment estimators. On generalizing the prescriptions outlined in (Schneider et al. 1998; Munshi & Coles 2003) to include the shear weights, as in Porth et al. (2020), one can work out expressions for the variance of the higher order direct estimators equation (16) for a given aperture. For the explicit derivation of for the variance of the third-order statistic, see Appendix D in the online supplementary material. From this analysis, we see that the general expression can be written

$$\sigma^2\left[\widehat{M^n_{\rm ap}}\right] = \sum_{\ell=0}^{n}\sum_{m=\ell}^{n} \frac{\sum_{\neq} w^2_{i_1}\cdots w^2_{i_m} w_{i_{m+1}}\cdots w_{i_n} w_{j_{m+1}}\cdots w_{j_n}}{\left(\sum_{\neq} w_{i_1}\cdots w_{i_n}\right)^2} \times C(n,\ell,m) M^{\ell}_{g,2} \left(\frac{\sigma^2_\epsilon}{2}\right)^{\ell} \left\langle \mathcal{M}^{m-\ell}_{s,2} \mathcal{M}^{2(n-m)}_{\rm ap} \right\rangle, \quad (35)$$

where the sum over the galaxy weights can again be decomposed as sums of (bivariate) power sums and the multiplicities are given by

$$C(n,\ell,m) = \frac{1}{\ell!(m-\ell)!} \left(\frac{n!}{(n-m)!}\right)^2. \quad (36)$$

For a discussion of the origin for the multiplicity factor $C(n, \ell, m)$, as well as a motivation of equation (35) and some of its limits we refer the reader to Appendix C; in particular, we obtain for the shape noise dominated limit

$$\sigma^2\left[\widehat{M^n_{\rm ap}}\right] \approx n! \frac{\sum_{\neq} w^2_{i_1}\cdots w^2_{i_n}}{\left(\sum_{\neq} w_{i_1}\cdots w_{i_n}\right)^2} \left(\frac{\sigma^2_\epsilon}{2}\right)^n M^n_{g,2}\,. \quad (37)$$

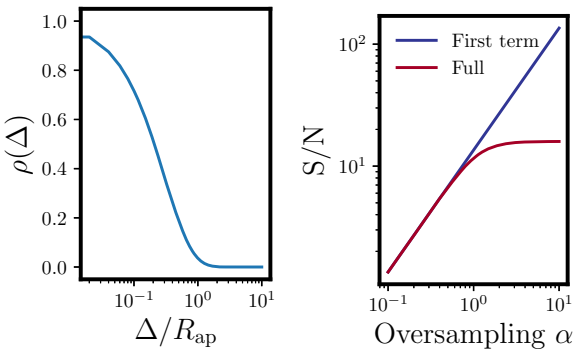

Figure 1. Example configuration of the correlation coefficient ρ (left-hand panel) and its effect on the signal contained in a survey field as predicted from equation (41) (right-hand panel).

The above formula gives the variance per aperture, thus for the estimator over the full survey field, equation (34), the variance can be written down as

$$\sigma^2\left[\hat{\mathfrak{M}}\right] = {\rm cov}\left(\frac{\sum_i w_i \hat{\mathfrak{M}}_i}{\sum_i w_i}, \frac{\sum_j w_j \hat{\mathfrak{M}}_j}{\sum_j w_j}\right) = S_2\, \sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right] + \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right]}{\left(\sum_i w_i\right)^2} \sum_{i\neq j} w_i w_j \rho(\hat{\mathfrak{M}}_i, \hat{\mathfrak{M}}_j)\,, \quad (38)$$

where in the above we have defined the cross-correlation coefficient between apertures whose centres are at position $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ to be:

$$\rho(\mathfrak{M}_i, \mathfrak{M}_j) \equiv \frac{\langle \mathfrak{M}_i, \mathfrak{M}_j \rangle}{\sqrt{\langle \mathfrak{M}_i, \mathfrak{M}_i \rangle \langle \mathfrak{M}_j, \mathfrak{M}_j \rangle}}\,. \quad (39)$$

Note that for the case of well-separated apertures, the cross-correlation coefficient will vanish and only the first summand needs to be taken into account, which for unity weights gives the familiar $1/N_{\rm ap}$ scaling of the variance. If the apertures are oversampled, this assumption is no longer valid and the term involving ρ must be included. Owing to the fact that ρ should only depend on the relative spatial distance Δ between the aperture centres, we can rewrite (38) as a weighted sum over all possible distances between aperture centre pairs:

$$\sigma^2\left[\hat{\mathfrak{M}}\right] = S_2\, \sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right] + \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right]}{\left(\sum_i w_i\right)^2} \sum_{b\in{\rm bins}} \left(\sum_{i,j\in\mathcal{I}_b} w_i w_j\right) \rho(\hat{\mathfrak{M}}, \Delta_b) \quad (40)$$

$$\approx \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right]}{N_{\rm ap}} + 2\pi \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right]}{A_{\rm survey}} \int_{R_{\rm ap}/\alpha}^{\infty} {\rm d}\Delta\Delta\rho(\hat{\mathfrak{M}}, \Delta), \quad (41)$$

where in the first step the bins are defined as a partition of the reals, and $\mathcal{I}_b \equiv \{i, j | \Delta(i, j) \in b\}$ collects all the aperture centre pairs falling into bin b. For the second step, we make the approximation that each aperture contains roughly the same signal such that the weights can be set to unity and we furthermore rewrote the expression in a continuous version, which makes the interpretation of the cross-term more concise. In particular, we parametrize the lower bound of the integral in terms of the aperture oversampling rate $\alpha \equiv R_{\rm ap}/\Delta_{\rm min}$.

In a realistic scenario, we expect ρ to to rapidly decrease from unity and then to approach zero for $\Delta \gg R_{\rm ap}$. An example of such a correlation coefficient is shown in Fig. 1. Here, we explicitly see

the importance of taking into account the cross-term once there is a substantial overlap between neighbouring apertures. In this example, we would infer that measuring the statistics with $\alpha \approx 2$ would be sufficient to extract most of the signal.

3.5 Implementation and scaling of the direct estimator

A practical implementation of equation (34) consists of three steps:

(i) Spatially organize the shape catalogue to allow for a fast assignment of galaxies to apertures.
(ii) For each aperture of the ensemble compute $\hat{\mathfrak{M}}_{\rm ap}$, the associated weight and optionally additional systematics (i.e. the coverage fraction c_k). Store each of these values in an array.
(iii) Based on some aperture selection and aperture weighting criteria $w_{\rm ap}$, update the weights and evaluate the weighted sum.

In what follows, we will explore each of these steps in more detail and for clarity, we will denote the number of galaxies in the survey and in the aperture as $N_{\rm g}$ and $N_{\rm g,\,ap}$, respectively.

3.5.1 Assigning galaxies to apertures

For our implementation, we use a spatial hashing data structure. We start by covering the survey footprint with an equal area mesh of $N_{\rm pix}$ pixels and create a hash table with the ID of each pixel as the key and the galaxy IDs as values. The hash function in our case is the ordinary pixel assignment function. For each aperture, we iterate over the associated galaxies within pixels that partially lie within the Q filter's support. The construction of the hash table scales as $\mathcal{O}\left(\max(N_{\rm pix}, N_{\rm g})\right)$ and the assignment is achieved in $\mathcal{O}\left(\max(N_{\rm pix,ap}, N_{\rm g,ap})\right)$ time per aperture. We found that when making a sensible choice of the mesh's coarseness, this data structure is more stable than a naive KD-tree based implementation as it does not require an additional range search operation which scales as $\mathcal{O}(\log(N_{\rm g}))$ per aperture and thus becomes a bottleneck for small apertures.

3.5.2 Computing the statistics per aperture

For the case of all radii being equal, we first compute the power sums in equations (22) and (23) and then recursively transform them to the corresponding moments via the recurrence relation (Comtet 1974)

$$B_{n+1}(x_1, \cdots, x_{n+1}) = \sum_{i=0}^{n} \binom{n}{i} B_n(x_1, \cdots, x_n) x_{i+1}, \tag{42}$$

where $B_0 \equiv 1$. Evaluating each power sum is linear in $N_{\rm g,\,ap}$ and for all practical applications the time taken for transforming to the $M_{\rm ap}^n$ basis can be neglected.

For the general case, we need to compute the relevant multivariable power sums equation (31) and bring them to the aperture moments basis by the transformation equation (32). In order to dynamically allocate and evaluate those expressions, we use a combinadic counting scheme to organize the power sum basis whereas the transformation equation is generated with the help of restricted growth strings (Knuth 2005).

3.5.3 Choice of weights for the averaging

Following our findings in Porth et al. (2020), we employ an inverse shot noise weighting scheme with an additional hard cutoff $c_{\rm lim}$ for the aperture coverage $c_{\rm ap}$, which for second-order statistics was found to lower the mask induced bias while increasing the signal-to-noise ratio compared to equal weights. The explicit form of the weights for the nth moment can be found from equation (37) when neglecting all constant contributions:

$$\mathcal{W}_{\rm ap}^{\rm (shot)}(c_{\rm lim}) \equiv \left[\frac{\sum_{(i_1,\cdots,i_n)} w_{i_1}^2 \cdots w_{i_n}^2}{\left(\sum_{(i_1,\cdots,i_n)} w_{i_1} \cdots w_{i_n}\right)^2}\right]^{-1} \mathcal{H}(c_{\rm ap} - c_{\rm lim})\,. \tag{43}$$

Dependent on whether we are dealing with the case of equal or unequal aperture radii the sums can be decomposed in a similar fashion as described above and evaluated together with the corresponding linearized direct estimator. As a further refinement one could also include the weights and completenesses of the surrounding apertures weighted by the spatial cross-correlation coefficient $\hat{\rho}$ – this would upweight apertures that are close to a mask as they cover more unique area.

4 RESULTS: APPLICATION TO GAUSSIAN MOCKS

4.1 Aperture mass statistics and Gaussian-lensing fields

In order to validate that our hierarchy of aperture mass moment estimators are unbiased and do indeed recover correct results, we first apply them to a set of Gaussian mock-lensing simulations. In this case, the whole moment hierarchy can be written as powers of the variance of the aperture mass. Hence, this motivates us to define the scaled aperture mass moments:

$$s_n(\vartheta_1) \equiv \frac{1}{(n-1)!!} \frac{\left\langle \mathcal{M}_{\rm ap}^n \right\rangle(\vartheta_1)}{\left[\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\vartheta_1)\right]^{n/2}} = \delta^K_{n,2\mathbb{N}}\,, \tag{44}$$

where the final equality is true for a Gaussian field only.

In order to test this, we generated 256 Gaussian lensing mocks. The methodology to create each mock was as follows:

(i) We first generate a Gaussian convergence field over a 12×12 deg^2 survey area. The area is tiled by a rectangular mesh of 8192^2 pixels. The variance of the convergence is obtained through specifying the convergence power spectrum, and we do this for a source distribution similar to that for the CFHTLenS survey (Fu et al. 2014).
(ii) We next obtain the shear field. This is done by Fourier transforming the convergence field and making use of the Kaiser & Squires (1993) approach.[7]
(iii) We then sample 4×10^6 galaxies into the survey footprint and use a multilinear interpolation of the shear field on to each galaxy.

Note that since we are assessing the accuracy of the estimators only we choose to set the intrinsic ellipticities of our source galaxies to zero. On repeating the analysis below when including this term we did not find a shift of the curves.

4.2 Computational scaling tests

Owing to the fact that each step of our algorithm is strictly linear, we expect a linear relationship between the elapsed time for estimator evaluation and the number of galaxies, for any given statistic. In

[7] In order to suppress edge effects introduced by the FFT we build the pixelated convergence field on an a plane having 16 times the area of the mock.

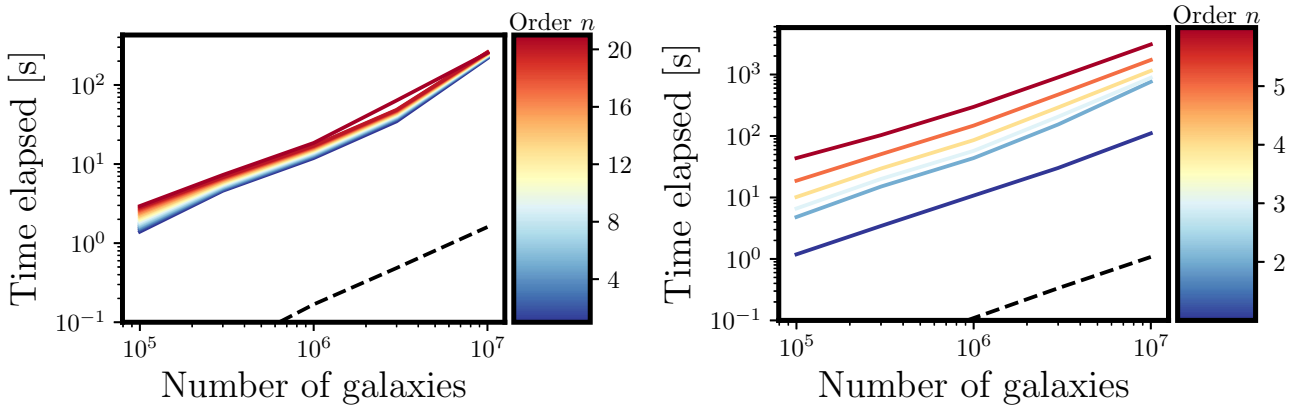

Figure 2. Computational complexity of the direct estimators for equal (left-hand panel) and unequal (right-hand panel) aperture radii as a function of the number of galaxies. All results are given for apertures of radius 10 arcmin which are oversampled by a factor of 16 ($\alpha = 4$) on a survey field of size $(12\text{deg})^2$. Different colours indicate different orders of the evaluated statistics. The black dashed line indicates the time spent in constructing the spatial hash. We see that for equal aperture radii the evaluation of higher order statistics basically comes for free, while for unequal radii there is a constant multiplicative offset based on the relative size of the radii and on the order which traces the number of multivariate power sums that need to be evaluated. All the scaling were obtained when running the estimator on a single CPU core.

addition, for the equal radius case, the order of the statistics should not strongly impact the evaluation time. However, for the unequal radius case, this does not necessarily hold true, since the computation depends on the relative sizes of the apertures as well as on the order of the statistics to be evaluated.

Fig. 2 shows the elapsed time of the direct estimator calculation for a Gaussian mock, where the number of sampled galaxies in the mock is increased. Focusing on the left-hand panel first, this shows the case for the standard aperture mass estimators with equal radii and here we compute all of the moments up to the 20th order. As expected the computational time for all of the moments scales linearly with the size of the problem and we also see that there is no obvious drop in performance for the higher order moments.

The right-hand panel of Fig. 2 is the same as the left-hand panel, but now for the case of unequal radii aperture mass moments, and here we only consider moments up to sixth order. There are two differences between the equal and non-equal radius case. First, we can see that there is a much larger multiplicative offset between adjacent orders for the generalized statistics. This is expected as the number of basis elements that need to be allocated in that case is given by 2^n compared to the n ones in the equal radius statistics. We also observe that for the second-order statistic the unequal radius calculation does roughly need four times as long as the equal radius one. We can explain this offset when noting that for our example the ratio of the largest and smallest scale was set to two. With our definition of the oversampling rate as being relative to the smallest aperture radius this implies that we need to allocate four times as many galaxies.

Finally, we note that the superior scaling of the direct estimator compared to traditional estimation methods should not come as a surprise. Looking back at the original definition equation (11) of the aperture mass one sees that it depends on the positions and shapes of the galaxies with respect to the aperture origin. In contrast, when switching to the description of aperture mass in terms of the shear correlation functions (i.e. Schneider et al. 2002a), the main dependence shifts to the relative distance and shapes between tuples of galaxies. This change of reference position makes the evaluation of correlation function based estimators intrinsically much more complex than a simple discretization of equation (11).

4.3 A hierarchy of aperture mass moments

Fig. 3 shows a comparison of the direct estimators for the second and fourth-order aperture mass moments as a function of angular scale as applied to the 256 Gaussian mocks. Here, we consider the case where all the aperture radii are equal (recall that for a Gaussian field all of the odd moments vanish). In both cases, the curves are in very good agreement with the Gaussian theory predictions, indicated by the solid red lines. We also note that for increasingly large aperture radii the measured results appear to be slightly below the theoretical expectation. This discrepancy can be attributed to finite field effects, as well as border effects being introduced by the Kaiser–Squires inversion method (see Pires et al. 2020, for a discussion).

Fig. 4 presents the measured s_n (see equation 44) for all of the aperture mass moments up to 10th order as a function of the aperture scale. We see that they are consistent with the Gaussian theoretical expectations. Note that in order to obtain this good agreement and circumvent the finite field effects described above, we used the ensemble mean of the measured aperture mass variance as the denominator in s_n.

Fig. 5 displays the fourth- and sixth-order multiscale aperture mass statistics as a function of the scale parameter. Note that there are a number of options for exploring the configuration dependence of the multiscale aperture mass moments, here we focus on fixing the ratio of the filter lengths and varying the overall scale of the configuration $\vec{a}$ with the parameter ϑ, e.g. for the kurtosis we would have

$$\left\langle \mathcal{M}_{\rm ap}^4 \right\rangle_{\rm c} \left(\vec{\vartheta}\right) \equiv \left\langle \mathcal{M}_{\rm ap}^4 \right\rangle_{\rm c} (a_1\vartheta, \cdots, a_4\vartheta), \tag{45}$$

where the constant $a_i \in \mathbb{R}_+$ specify the configuration. The estimates shown in the figure were obtained using our generalized estimator equation (32). As for the previous cases, we find good agreement between the measurements and the Gaussian predictions, which were

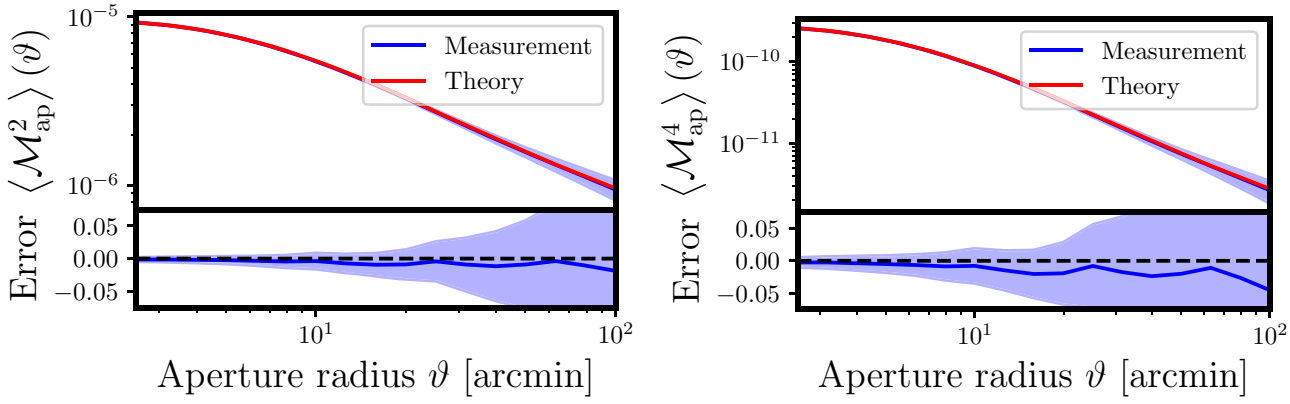

Figure 3. Comparison of the measured aperture mass moments with their theoretical prediction. Left-hand panel: The upper subpanel shows the aperture mass dispersion as a function of the aperture scale. The red line shows the theoretical predictions evaluated from the input power spectrum and the blue line shows the measurement from the mocks. The blue shaded regions show the standard deviation of the corresponding measurement across the ensemble. The lower subpanel shows the relative error between the measurement and the theory, with the line styles as before. Right-hand panel**:** Same as left-hand panel, but for the kurtosis of aperture mass.

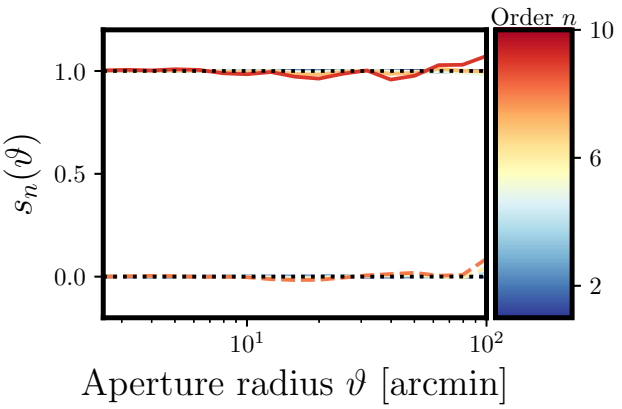

Figure 4. Scaled nth-order aperture mass moments $s_n(\vartheta)$ (see Equation 44), measured in the ensemble of 256 Gaussian mocks, as a function of the aperture scale, for all moments up to 10th order. The solid lines of varying colours show the mean of the measurements. The dotted black lines show the Gaussian theoretical expectations. For a Gaussian mock, the even order $s_n(\vartheta)$ give unity, and the odd ones vanish.

obtained by making use of equation (13) and Wicks theorem for the convergence polyspectra (Bernardeau et al. 2002).

5 RESULTS: DETECTION SIGNIFICANCE OF HIGHER ORDER MOMENTS

In this section, we now turn to the question of the detection significance of higher order aperture statistics from current and future surveys.

5.1 The SLICS mocks

In order to answer this question, we make use of the SLICS[8] mocks – this is a large suite of lensing mock catalogues generated from a large set of cosmological N-body simulations (for full details see Harnois-Déraps et al. 2018). Each SLICS mock corresponds to a survey area of 100 deg^2. These are generated from the past light cone extracted from fully independent gravity-only N-body simulations, which evolve $N = 1532^3$ particles within a comoving box of length $L = 505\ h^{-1}$Mpc. The lensing maps are constructed using the Born approximation. We adopt the catalogues for which the galaxies are randomly distributed within the lightcone according to the KiDS-450 source distribution (Hildebrandt et al. 2017). The shape noise has been set to $\sigma_\epsilon = 0.29$ per shear component. In order to mimic a constraining power that is comparable to the KiDS-1000 data while not being too noisy, we rescale the errorbars by a factor of $\sqrt{10}$. This provides us with effectively 819 simulated 1000(deg)^2 surveys with which to perform our analysis.

When estimating the aperture mass statistics from the SLICS mocks using the estimator given by equation (34), the achievable signal-to-noise ratio ratio will depend on the number of sampled apertures selected. If too few are chosen then our estimate will be inefficient, on the other hand, due to the fact that there are aperture-to-aperture correlations choosing too many will capture all of the available information, but ultimately will be computationally inefficient. We therefore expect that the information will saturate for a given oversampling rate, and that to sample at a higher rate would be of little use. To investigate this, we proceed as in Porth et al. (2020) and place apertures on a regular grid with spacing Δ, corresponding to an aperture oversampling rate of $\alpha \equiv \min(\{\vartheta_i\})/2\Delta$.

5.2 Measurement in the SLICS mocks

Fig. 6 shows the detection significance of the equal radii aperture mass statistics for the second-, third-, fourth- and fifth-order aperture mass statistics as a function of the aperture scale and for various choices of the oversampling rate. For the second-order statistics, we also plot the theoretical prediction of the aperture mass dispersion evaluated from equation (11), where the convergence power spectrum was computed with CCL[9] (Chisari et al. 2019) using `Halofit` (Smith et al. 2003), but with the modifications of Takahashi et al. (2012), as the matter power spectrum. While part of the difference between the curves for small aperture radii could be attributed to uncertainties in `Halofit`, our suspicion is

[8]https://slics.roe.ac.uk/

[9]https://github.com/LSSTDESC/CCL

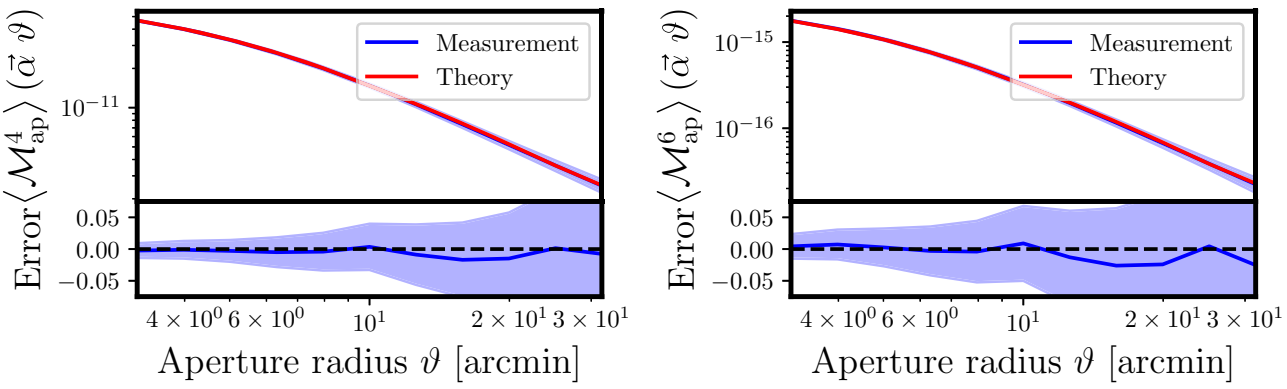

Figure 5. Multiscale aperture mass moments as a function of the scale parameter ϑ, measured in the ensemble of 256 Gaussian mocks. Line styles are the same as in Fig. 3. Left-hand panel: the fourth-order aperture statistics. In this case, the vector of aperture scales was set to $\boldsymbol{\alpha}$=(0.5,0.8,1.,2.). Right-hand side: Same as left-hand side, but this time for the sixth-order statistic with the vector of aperture scales set to $\boldsymbol{\alpha}$ = (0.5,0.7,1.,1.1,1.5,2.).

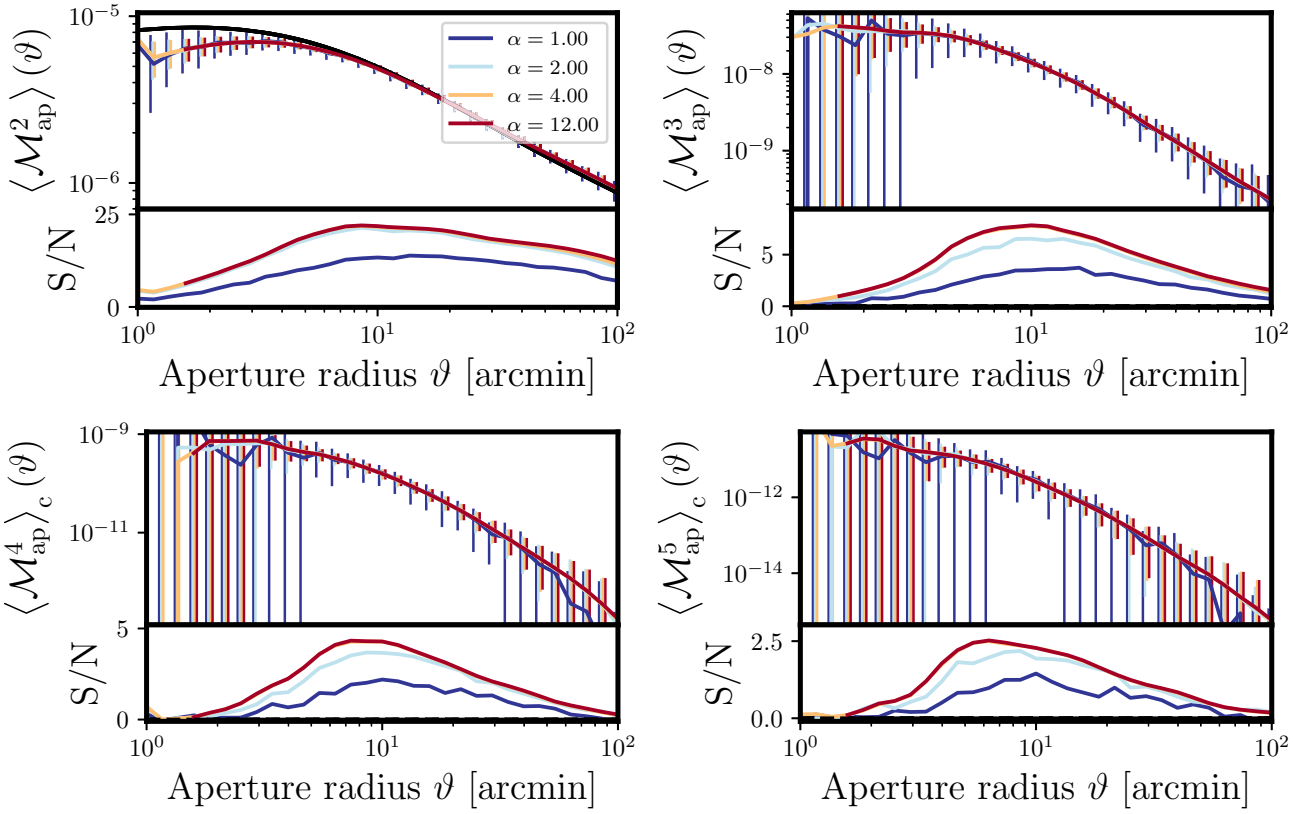

Figure 6. Measurement of the aperture mass statistics in the SLICS simulation suite for different aperture oversampling rates α. All measurements were done on an ensemble of 819 realizations with an angular area of 100 deg^2 each, where the $n(z)$ follows the KiDS-450 distribution. The upper part of the panels corresponds to the mean and rescaled standard deviation from the ensembles. The lower panel shows the signal-to-noise ratio for the corresponding statistics when rescaled to match a 1000 deg^2 survey. For the aperture mass dispersion, we additionally plot the theoretical prediction as the black line. For the fourth- and fifth-order plots, we restrict ourselves to the contribution of the connected part of the convergence polyspectra. We see that choosing an oversampling parameter of $\alpha \gtrsim 4$ recovers most of the information.

that they mostly stem from the limited particle mass resolution in the SLICS mocks (see fig. 6 in Harnois-Déraps et al. 2018 for the resulting suppression of the shear correlation functions for small separations).

Several important points are worth noting from these measurements. First, we see that for a KiDS-1000 like survey there is sufficient fidelity to detect the aperture mass statistics up to fourth order,[10] with the signal-to-noise ratio peaking at an aperture size of around $\vartheta \approx 10$ arcmin for all statistics. This is exciting, as this has never before been achieved with standard correlation function based

[10]We find the cumulative detection significance of the fifth-order statistics to be at the 2.9σ level.

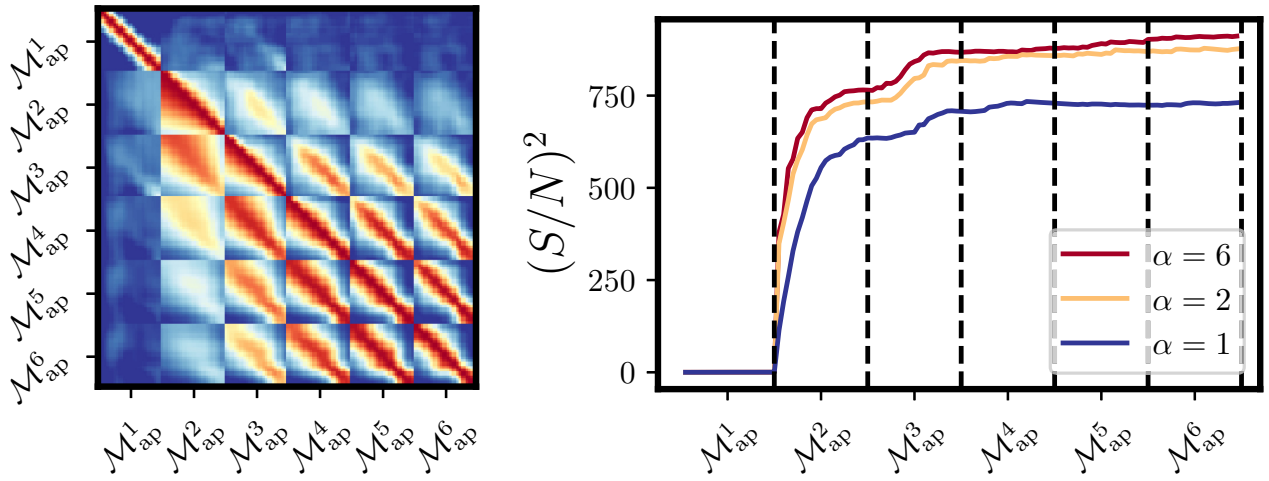

Figure 7. Correlation coefficient matrix (left-hand panel) and cumulative detection significance (right-hand panel) for connected moments of the aperture mass statistics. We take into account aperture sizes between 10 and 100 arcmin. In the correlation matrix, the lower triangle shows the results without shape noise while the upper part includes this term and serves as the basis for the computation of the detection significance.

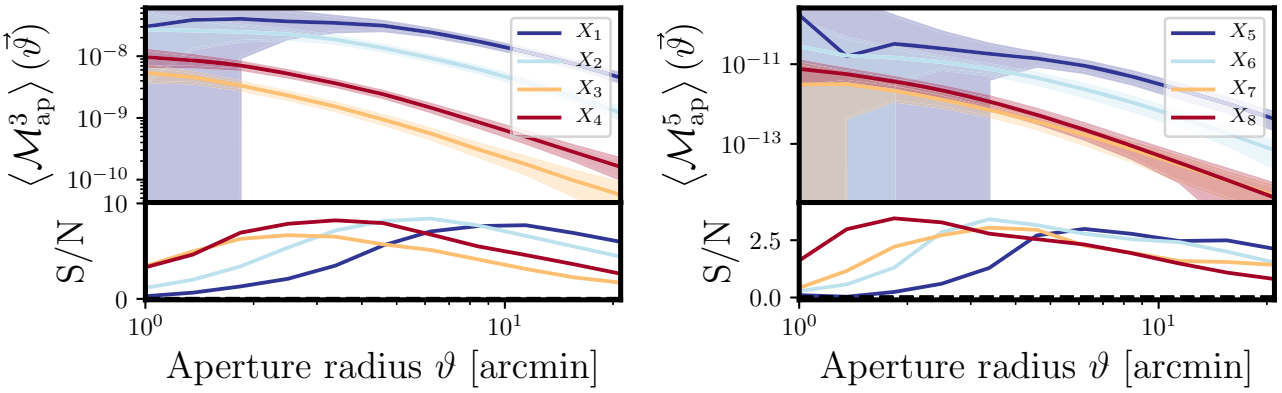

Figure 8. Measurements of the unequal radii aperture mass statistics of third (left-hand panel) and fifth (right-hand panel) order in the SLICS simulation suite. Each line corresponds to a different set of relative aperture sizes as given in Table 1.

estimators, and if correct would represent the first robust detection of these statistics using these methods. Second, while for the case of the two-point statistics the signal-to-noise ratio ratio (shown in lower sub-panels for each plot) falls off slowly for larger apertures, this ratio approaches zero faster for the connected parts of the higher order statistics.[11] Third, while an aperture oversampling rate of $\alpha \approx 2$ seems sufficient to capture all the signal for second-order statistics, it misses some information for subsequent orders where it becomes necessary to use $\alpha \gtrsim 4$.

Fig. 7 displays the correlation structure of the aperture mass cumulants as well as the cumulative detection significance. We only consider measurements with $\vartheta \geq 10$ arcmin as this is where the SLICS mocks do agree reasonably well with the theoretical predictions and and due to the fact that the robust theoretical modelling of those statistics might reach its limits at around those scales. We see that while for shape noise free ellipticity catalogues there are strong correlations for small aperture radii, this is not the case for the realistic mocks in which those scales are still shape noise dominated.

[11]Owing to the fact that the aperture mass has zero mean, the full and connected moments differ only for even order moments of four or more.

Table 1. Cofigurations of the aperture radii displayed in Fig. 8.

Third order		Fifth order	
Label	Configuration $\vec{a}$	Label	Configuration $\vec{a}$
X_1	(1, 1, 1)	X_5	(1, 1, 1, 1, 1)
X_2	(1, 2, 2)	X_6	(1, 1, 2, 2, 2)
X_3	(1, 5, 5)	X_7	(1, 1, 5, 5, 5)
X_4	(1, 3, 5)	X_8	(1, 2, 3, 4, 5)

We further note large correlations around the diagonal between different orders, where the degree of correlation increases with the order of the cumulants. For the cumulative detection significance, we see that that the cumulants beyond third order do not add a substantial amplitude to the cumulative signal-to-noise ratio. This is expected, given the relatively lower signal-to-noise ratio as well as the larger portion of cross-covariances that need to be taken into account. One should note that this type of analysis does not imply that the higher order cumulants are obsolete as they still may add complementary information by breaking cosmological parameter degeneracies.

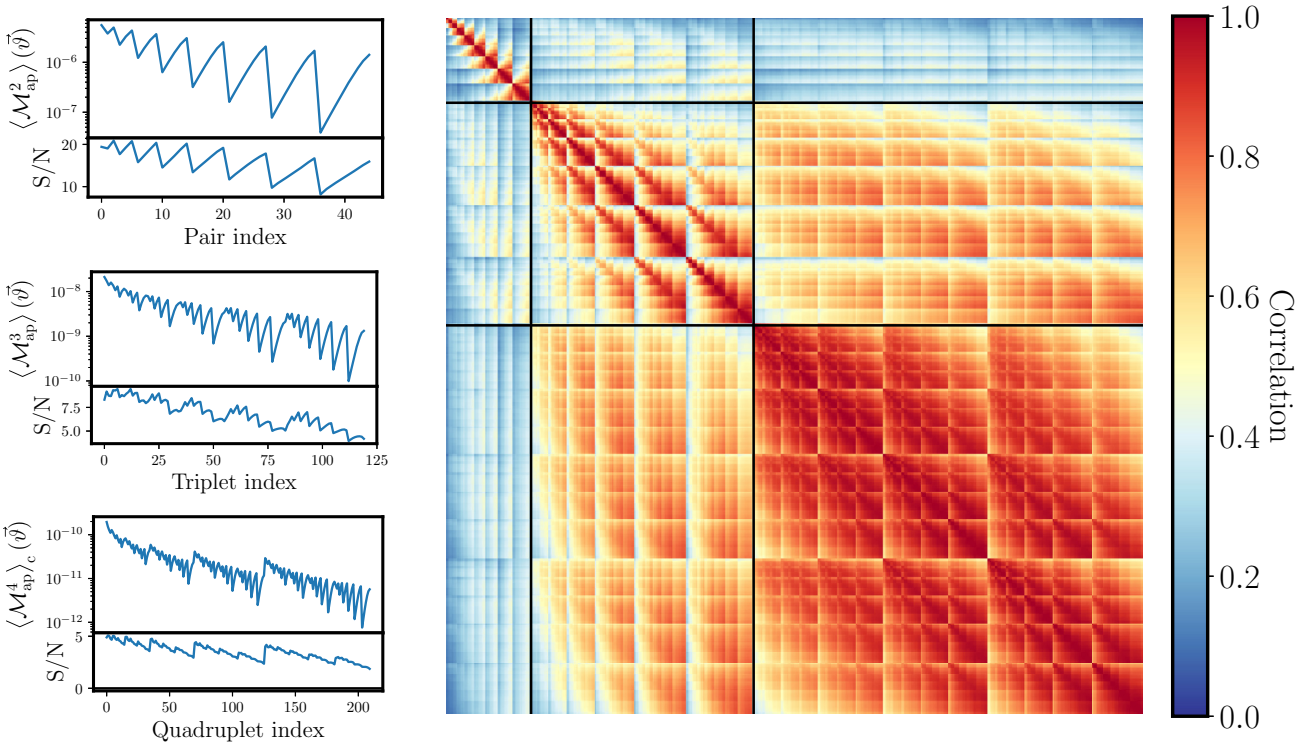

Figure 9. Measurements (left-hand panel) and correlation matrix (right-hand panel) of the multiscale aperture mass statistics of second, third and fourth order. For each of those statistics, we compute all configurations for 10 logarithmically spaced radii between 5 and 50 arcmin in which all the apertures have unequal radii. The black lines indicate the blocks of the (cross-) correlations of different orders.

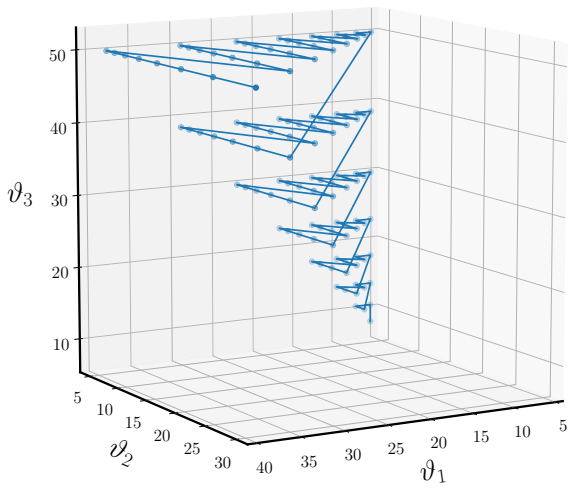

Figure 10. Path in which the set of non-redundant aperture scales for the third-order statistics is traversed. The starting point is the lower corner. The subpath in the $\vartheta_3 = 50$ arcmin plane corresponds to the full path taken for the second-order statistics.

5.3 Multiscale aperture mass measurements

We now shift to the measurement of the multiscale statistics for which there are a number of ways on how to select various aperture scale multiplets. In Fig. 8, we focus on a fixed set of aperture propositions and then simply scale them with a single parameter ϑ. The different configurations $\vec{a}$ of aperture radii that we have employed are shown in Table 1. We see that for both, the third- and the fifth-order moments there does not appear to be a strong decline in detection significance for multiscale apertures compared to the associated moments, even if the relative spread of radii is large.

Another way to select aperture scale multiplets for a statistic of order n is choose a list of $m \geq n$ aperture scales and to compute the statistics for each choice of n elements within that list. For our purposes, we choose the subset in which none of the aperture radii are equal, as this speeds up our calculation, see Appendix B3 for the details. On the left-hand side of Fig. 9, we show our measurements for the second-, third- and fourth-order connected cumulants of the multiscale aperture mass statistic using ten logarithmically spaced scales between 5 and 50 arcmin. The first index of the multiplet corresponds to the selection of the smallest possible aperture scales from which we then start choosing the next lowest radius in the subsequent dimension up until we reach the combination of the largest possible set of aperture radii – for an example of this path for the third-order statistics see Fig. 10. Recalling that the second-order aperture mass statistic is simply a filtered version of the power spectrum, we should not expect the multiscale extension add any information to that order.[12] For the three statistics, we again find a detection significance that is comparable to the equal scale case, meaning that we can extract substantial signal from convergence spectra configurations which are not corresponding to regular polygons. On the right-hand side of Fig. 9, we plot the joint correlation coefficient of the multiscale cumulants. On the investigated range of scales, we only find a slight to modest correlation between the higher order multiscale statistics and the

[12]For a different form of the Q filter function like the one proposed in Crittenden et al. (2001) one can easily work this out analytically, see.i.e. Schneider et al. (2005).

second-order one. It also appears that the higher order cumulants exhibit a stronger autocorrelation and cross-correlation. However, this is (at least partially) an artefact of the range and sampling density of the chosen radii.

6 CONCLUSIONS AND DISCUSSION

In this paper, we have explored an alternative method for estimating the aperture mass statistics in weak-lensing cosmic shear surveys. This study extended our previous work (Porth et al. 2020) in a number of ways: First, we generalized the direct estimator approach to higher statistics, and showed how to rewrite the standard estimator as a product of linear order time sums. Second, we provided the details of the computation of the variance of these estimators. Third, we further generalized the aperture mass statistics to include the multiscale approach. Again, we showed how one could estimate these using linear order products of power sums. The work can be summarized as follows:

In Section 2, we reviewed the background theory of cosmological weak lensing and showed how the connected cumulants of the aperture mass statistics are related to the convergence polyspectra.

In Section 3, we introduced the direct estimator for moments of the aperture mass statistics. We then gave expressions for how the nested sums can be decomposed into a linear combination of products of (multivariate) power sums that facilitates a linearly scaling estimation procedure in the number of galaxies within an aperture. We then generalized this estimator to an ensemble of overlapping apertures and computed its variance. We argued that the aperture cross-correlation coefficient leads to a substantial correction to the naive $1/N$ scaling if the apertures are not well separated, and that it also can be used to assess the degree of aperture oversampling that is necessary to capture most of the available information. Finally, we gave a detailed explanation of the algorithms used for our implementation.

In Section 4, we successfully validated our method on Gaussian mock simulations and furthermore verified the linear scaling.

In Section 5, we turned to the SLICS simulation suite and assessed the signal-to-noise ratio of the statistics for a 1000° survey following a KiDS-450 like $n(z)$ distribution function. We found that with these specifications significant detections of up to fourth order can be expected for the equal and unequal radii cumulants and that an aperture oversampling rate of at least four extracts nearly all the signal.

In this paper, we have neglected the impact of survey masks on the measurement process and the possible bias that this could induce, the exploration of this is sufficient to warrant its own publication and this is the subject of our associated publication (Porth et al., in preparation). Throughout this paper, we were mainly concerned with making the extraction of information from higher order statistics of galaxy shape catalogues computationally feasible and accurate. However, we remained agnostic about further challenges that need to be addressed before applying our methods to real data. For example, one should investigate the required PSF modelling, shape measurement and shear bias calibration quality to not introduce substantial biases in the measurement. Additionally, the range of measurements that can ultimately be used for obtaining cosmological parameter constraints will be limited to the scales for which one can theoretically accurately model those higher order statistics.

ACKNOWLEDGEMENTS

We would like to thank Patrick Simon, Laura Marian, Stefan Hilbert, Peter Schneider, Cora Uhlemann, and Gary Bernstein for useful discussions, as well as the anonymous referee for helpful comments. We would like to thank Joachim Harnois-Deraps for making public the SLICS mock data, which can be found at http://slics.roe.ac.uk/. LP acknowledges sup port from a STFC Research Training Grant (grant number ST/R505146/1). RES acknowledges support from the STFC (grant number ST/P000525/1, ST/T000473/1). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National eInfrastructure. This research made use of numpy, a library used for scientific computing and technical computing and MATPLOTLIB, a Python library for publication quality graphics (Hunter 2007; Harris et al. 2020).

DATA AVAILABILITY

The SLICS mock catalogues are available at https://slics.roe.ac.uk/. Additional data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Aihara H. et al., 2018, Publ. Astron. Soc. Japan, 70, S4
Asgari M. et al., 2021, A&A, 645, A104
Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Barthelemy A., Codis S., Bernardeau F., 2020, MNRAS, 50, 5204
Bernardeau F., Valageas P., 2000, A&A, 364, 1
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Blandford R. D., Saust A. B., Brainerd T. G., Villumsen J. V., 1991, MNRAS, 251, 600
Byun J., Eggemeier A., Regan D., Seery D., Smith R. E., 2017, MNRAS, 471, 1581
Chisari N. E. et al., 2019, ApJS, 242, 2
Comtet L., 1974, Advanced Combinatorics: The Art of Finite and Infinite Expansions. Springer, Heidelberg
Crittenden R. G., Natarajan P., Pen U.-L., Theuns T., 2001, ApJ, 559, 552
Dodelson S., 2003, Modern cosmology. Academic Press, San Diego, CA
Dodelson S., 2017, Gravitational Lensing. Cambridge Univ. Press, Cambridge
Fabbian G., Calabrese M., Carbone C., 2018, J. Cosmol. Astropart. Phys., 2018, 050
Friedrich O., Seitz S., Eifler T. F., Gruen D., 2016, MNRAS, 456, 2662
Fu L. et al., 2014, MNRAS, 441, 2725
Harnois-Déraps J. et al., 2018, MNRAS, 481, 1337
Harris C. R. et al., 2020, Nature, 585, 357
Heydenreich S., Brück B., Harnois-Déraps J., 2020, A&A, 648, 18
Hikage C. et al., 2019, Publ. Astron. Soc. Japan, 71, 43
Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A&A, 499, 31
Hilbert S., Marian L., Smith R. E., Desjacques V., 2012, MNRAS, 426, 2870
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jain B., Seljak U., 1997, ApJ, 484, 560
Jarvis M., Bernstein G. M., Fischer P., Smith D., Jain B., Tyson J. A., Wittman D., 2003, AJ, 125, 1014
Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
Kacprzak T. et al., 2016, MNRAS, 463, 3653
Kaiser N., 1995, ApJ, 439, L1
Kaiser N., 1998, ApJ, 498, 26
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kaiser N., Wilson G., Luppino G. A., 2000, preprint (astro-ph/0003338)
Kayo I., Takada M., Jain B., 2013, MNRAS, 429, 344

Kilbinger M., 2015, Rep. Prog. Phys., 78, 086901
Kilbinger M., Schneider P., 2005, A&A, 442, 69
Knuth D. E., 2005, The Art of Computer Programming, Volume 4, Fascicle 3: Generating All Combinations and Partitions. Addison-Wesley Professional, Boston, MA
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
LSST, 2009, preprint (arXiv:0912.0201)
Mandelbaum R., 2018, ARA&A, 56, 393
Marian L., Smith R. E., Hilbert S., Schneider P., 2012, MNRAS, 423, 1711
Marian L., Smith R. E., Hilbert S., Schneider P., 2013, MNRAS, 432, 1338
Martinet N., Harnois-Déraps J., Jullo E., Schneider P., 2021, A&A, 646, 16
Massey R. et al., 2013, MNRAS, 429, 661
Miralda-Escude J., 1991, ApJ, 380, 1
Munshi D., Coles P., 2003, MNRAS, 338, 846
Munshi D., Valageas P., 2005, MNRAS, 356, 439
Munshi D., Valageas P., Barber A. J., 2004, MNRAS, 350, 77
Pires S. et al., 2020, A&A, 638, A141
Porth L., Smith R. E., Simon P., Marian L., Hilbert S., 2020, MNRAS, 499, 2474
Pratten G., Lewis A., 2016, J. Cosmol. Astropart. Phys., 2016, 047
Sato M., Takada M., Hamana T., Matsubara T., 2011, ApJ, 734, 76
Schneider P., 1996, MNRAS, 283, 837
Schneider P., 1998, ApJ, 498, 43
Schneider P., 2006a, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C. S., Wambsganss J., eds, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro. Springer-Verlag, Berlin, p. 1
Schneider P., 2006b, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C. S., Wambsganss J., eds, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro. Springer-Verlage, Berlin, p. 269
Schneider P., Kilbinger M., 2007, A&A, 462, 841
Schneider P., Lombardi M., 2003, A&A, 397, 809
Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, MNRAS, 296, 873
Schneider P., van Waerbeke L., Mellier Y., 2002a, A&A, 389, 729
Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002b, A&A, 396, 1
Schneider P., Kilbinger M., Lombardi M., 2005, A&A, 431, 9
Schneider P., Eifler T., Krause E., 2010, A&A, 520, A116
Scoccimarro R., Frieman J., 1996, ApJS, 105, 37
Seitz S., Schneider P., Ehlers J., 1994, Classi. Quantum Gravity, 11, 2345
Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020
Smith R. E. et al., 2003, MNRAS, 341, 1311
Szapudi I., Szalay A. S., 1997, ApJ, 481, L1
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Troxel M. A., Ishak M., 2015, Phys. Rep., 558, 1
Troxel M. A. et al., 2018, Phys. Rev. D, 98, 043528
Van Waerbeke L. et al., 2000, A&A, 358, 30
Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nature, 405, 143
Zhang P., Liguori M., Bean R., Dodelson S., 2007, Phys. Rev. Lett., 99, 141302

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Appendix D. Variance of the Direct Estimator for the Aperture Mass Skewness.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: DERIVATIONS OF APERTURE MASS SKEWNESS AND KURTOSIS ESTIMATORS

In the following, we will derive the accelerated direct estimators for the third- and fourth-order aperture mass moments. For properly treating summation indices, we add to our notation (17) the following generalizations that deal with individual indices being set equal with each other:

$$\sum_{\substack{i_1,\cdots,i_l\cdots i_{m-1},i_{m+1},\cdots,i_n \\ i_l=i_m}} \equiv \sum_{i_1}\cdots\sum_{i_{m-1}}\sum_{i_{m+1}}\cdots\sum_{i_n}, \tag{A1}$$

$$\sum_{\substack{(i_1,\cdots,i_l\cdots i_{m-1},i_{m+1}\cdots i_n) \\ i_l=i_m}} \equiv \sum_{i_1}\sum_{i_2\neq i_1}\cdots\sum_{i_{m-1}\neq\cdots\neq i_1}\sum_{i_{m+1}\neq\cdots\neq i_1}\cdots\sum_{i_n\neq\cdots\neq i_1}. \tag{A2}$$

A1 Derivation of the estimator for $\widehat{M_{\rm ap}^3}$

Let us compute the derivation of the skewness $\widehat{M_{\rm ap}^3}$ of the aperture mass. The standard direct estimator is given by

$$\widehat{M_{\rm ap}^3} = \left(\pi\vartheta^2\right)^3 \frac{\sum_{(i,j,k)}^N w_i w_j w_k Q_i Q_j Q_k e_{t,i} e_{t,j} e_{t,k}}{\sum_{(i,j,k)}^N w_i w_j w_k}. \tag{A3}$$

It can be shown using the methods described in Schneider et al. (1998) and Porth et al. (2020) that this leads to an unbiased estimator of the skewness. We can rewrite the above estimator by noting that an unconstrained triple sum can be decomposed into the following partial sums:

$$\sum_{i,j,k}^N = \sum_{(i,j,k)}^N + \sum_{\substack{(i,j)\\ i=k}}^N + \sum_{\substack{(i,j)\\ j=k}}^N + \sum_{\substack{(i,k)\\ i=j}}^N + \sum_{i=j=k}^N. \tag{A4}$$

This can be rearranged to give

$$\sum_{(i,j,k)}^N = \sum_{i,j,k}^N - \sum_{\substack{(i,j)\\ i=k}}^N - \sum_{\substack{(i,j)\\ j=k}}^N - \sum_{\substack{(i,k)\\ i=j}}^N - \sum_{i=j=k}^N. \tag{A5}$$

Similarly, the unconstrained double sum can be decomposed and rearranged in the following manner:

$$\sum_{i,j} = \sum_{(i,j)} + \sum_{i=j} \;\Rightarrow\; \sum_{(i,j)} = \sum_{i,j} - \sum_{i=j}. \tag{A6}$$

Using this result repeatedly in equation (A5) allows us to rewrite the constrained sums as unconstrained sums:

$$\sum_{(i,j,k)}^{N} = \sum_{i,j,k}^{N} - \left(\sum_{\substack{i,j\\ i=k}}^{N} - \sum_{i=j=k}^{N}\right) - \left(\sum_{\substack{i,j\\ j=k}}^{N} - \sum_{i=j=k}^{N}\right) - \left(\sum_{\substack{i,k\\ i=j}}^{N} - \sum_{i=j=k}^{N}\right) - \sum_{i=j=k}^{N}$$
$$= \sum_{i,j,k}^{N} - \sum_{\substack{i,j\\ i=k}}^{N} - \sum_{\substack{i,j\\ j=k}}^{N} - \sum_{\substack{i,k\\ i=j}}^{N} + 2\sum_{i=j=k}^{N} = \sum_{i,j,k}^{N}\left[1 - \delta^K_{j,k} - \delta^K_{k,i} - \delta^K_{i,j} + 2\delta^K_{i,j}\delta^K_{i,k}\right] . \tag{A7}$$

Hence, on repeatedly using this result we can rewrite the sum in the numerator and denominator of equation (A3) to give us an alternate form for the skewness as:

$$\widehat{M^3_{\rm ap}} = \left(\pi\vartheta^2\right)^3 \frac{\left[\sum_{i,j,k}^{N} w_i w_j w_k Q_i Q_j Q_k e_{t,i} e_{t,j} e_{t,k} - 3\sum_{i,j}^{N} w_i w_j^2 Q_i Q_j^2 e_{t,i} e^2_{t,j} + 2\sum_{i}^{N} w_i^3 Q_i^3 e^3_{t,i}\right]}{\left[\sum_{i,j,k}^{N} w_i w_j w_k - 3\sum_{i,j}^{N} w_i w_j^2 + 2\sum_{i}^{N} w_i^3\right]}. \tag{A8}$$

If we now divide through each term by $(\sum_i^N w_i)^3$ and recall expressions equations (22) and (23), we see that our estimator becomes

$$\widehat{M^3_{\rm ap}} = \frac{M^3_{\rm s,1} - 3M_{\rm s,2}M_{\rm s,1} + 2M_{\rm s,3}}{1 - 3S_2 + 2S_3}. \tag{A9}$$

A2 Derivation of the estimator for $\widehat{M^4_{\rm ap}}$

The standard direct estimator for the kurtosis of aperture mass is given by

$$\widehat{M^4_{\rm ap}} = \left(\pi\vartheta^2\right)^4 \frac{\sum_{(i,j,k,l)}^{N} w_i w_j w_k w_l Q_i Q_j Q_k Q_l e_{t,i} e_{t,j} e_{t,k} e_{t,l}}{\sum_{(i,j,k,l)}^{N} w_i w_j w_k w_l} . \tag{A10}$$

We follow similar steps to the derivation of the skewness and note that the unconstrained quadruple sum can be written as

$$\sum_{i,j,k,l}^{N} = \sum_{(i,j,k,l)}^{N} + \left[\sum_{\substack{(i,j,k)\\ i=l}}^{N} + 5\text{ perms}\right] + \left[\sum_{\substack{(i,j)\\ i=k,j=l}}^{N} + \sum_{\substack{(i,j)\\ i=l,j=k}}^{N} + \sum_{\substack{(i,k)\\ i=j,k=l}}^{N}\right] + \left[\sum_{\substack{(i,j)\\ j=k=l}}^{N} + \sum_{\substack{(i,l)\\ i=j=k}}^{N} + \sum_{\substack{(i,k)\\ i=j=l}}^{N} + \sum_{\substack{(i,j)\\ i=k=l}}^{N}\right] + \sum_{i=j=k=l}^{N} , \tag{A11}$$

which on rearranging leads us to

$$\sum_{(i,j,k,l)}^{N} = \sum_{i,j,k,l}^{N} - \left[\sum_{\substack{(i,j,k)\\ i=l}}^{N} + 5\text{ perms}\right] - \left[\sum_{\substack{(i,j)\\ i=k,j=l}}^{N} + \sum_{\substack{(i,j)\\ i=l,j=k}}^{N} + \sum_{\substack{(i,k)\\ i=j,k=l}}^{N}\right] - \left[\sum_{\substack{(i,j)\\ j=k=l}}^{N} + \sum_{\substack{(i,l)\\ i=j=k}}^{N} + \sum_{\substack{(i,k)\\ i=j=l}}^{N} + \sum_{\substack{(i,j)\\ i=k=l}}^{N}\right] - \sum_{i=j=k=l}^{N} . \tag{A12}$$

We now make use of our previous results to rewrite the constrained sums, on the right-hand side, of the expression as unconstrained sums:

$$\sum_{(i,j,k,l)}^{N} = \sum_{i,j,k,l}^{N} - \left\{\left[\sum_{\substack{i,j,k\\ i=l}}^{N} - \sum_{\substack{i,j\\ i=k=l}}^{N} - \sum_{\substack{i,j\\ j=k,i=l}}^{N} - \sum_{\substack{i,k\\ i=j=l}}^{N} + 2\sum_{i=j=k=l}^{N}\right] + 5\text{ perms}\right\} - \left\{\left[\sum_{\substack{i,j\\ i=k,j=l}}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{\substack{i,j\\ i=l,j=k}}^{N} - \sum_{i=j=k=l}^{N}\right]\right.$$
$$\left. + \left[\sum_{\substack{i,k\\ i=j,k=l}}^{N} - \sum_{i=j=k=l}^{N}\right]\right\} - \left\{\left[\sum_{\substack{i,j\\ j=k=l}}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{\substack{i,l\\ i=j=k}}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{\substack{i,k\\ i=j=l}}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{\substack{i,j\\ i=k=l}}^{N} - \sum_{i=j=k=l}^{N}\right]\right\} - \sum_{i=j=k=l}^{N} . \tag{A13}$$

On making repeated use of the Kroneker delta symbol this can now be compactly written as

$$\sum_{(i,j,k,l)}^{N} = \sum_{i,j,k,l}^{N}\left[1 - \left\{\left[\delta^K_{k,l} - \delta^K_{j,k}\delta^K_{k,l} - \delta^K_{i,k}\delta^K_{k,l} - \delta^K_{i,j}\delta^K_{k,l} + 2\delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + 5\text{ perms}\right\} - \left\{\left[\delta^K_{i,j}\delta^K_{k,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\right.$$
$$+ \left[\delta^K_{i,k}\delta^K_{j,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{i,l}\delta^K_{j,k} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\left.\right\} - \left\{\left[\delta^K_{i,j}\delta^K_{i,k} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\right.$$
$$\left. + \left[\delta^K_{i,j}\delta^K_{i,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{i,k}\delta^K_{i,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{j,k}\delta^K_{j,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\right\} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\Bigg] . \tag{A14}$$

On collecting, cancelling, and grouping like terms, we see that this can be written as

$$\sum_{(i,j,k,l)}^{N} = \sum_{i,j,k,l}^{N} \Big[1 - \left(\delta^K_{i,j} + \delta^K_{i,k} + \delta^K_{i,l} + \delta^K_{j,k} + \delta^K_{j,l} + \delta^K_{k,l}\right) + \left\{\delta^K_{i,j}\delta^K_{k,l} + \delta^K_{i,k}\delta^K_{j,l} + \delta^K_{i,l}\delta^K_{j,k}\right\} + \left\{\delta^K_{i,j}\delta^K_{i,k} + \delta^K_{i,j}\delta^K_{i,l} + \delta^K_{i,k}\delta^K_{i,l} + \delta^K_{j,i}\delta^K_{j,k} + \delta^K_{j,i}\delta^K_{j,l} + \delta^K_{j,k}\delta^K_{j,l} + \delta^K_{k,i}\delta^K_{k,l} + \delta^K_{k,j}\delta^K_{k,l}\right\} - 6\delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\Big]. \tag{A15}$$

Hence, on making repeated use of equation (A15) in equation (A10) and along with equations (22) and (23), the estimator for kurtosis of aperture mass becomes

$$\widehat{M^4_{\rm ap}} = \frac{M^4_{\rm s,1} - 6M_{\rm s,2}M^2_{\rm s,1} + 3M^2_{\rm s,2} + 8M_{\rm s,3}M_{\rm s,1} - 6M_{\rm s,4}}{1 - 6S_2 + 3S_2^2 + 8S_3 - 6S_4}. \tag{A16}$$

APPENDIX B: A PROOF OF THE GENERAL THEOREM FOR ARBITRARY ORDER APERTURE MASS STATISTICS

In this section, we provide a derivation of the general form of the n-point aperture mass statistic estimator given by equation (24). At the time of writing, we are not aware that the combinatoric methods that we have used in the derivation of the general expression have been used before in the cosmological context, and therefore provide a brief overview of them – in particular the Bell polynomials. In what follows, we will try to not rely on advanced mathematical methods, but instead use a basic framework to explain how the Bell polynomials are linked to set partitions, and finally how they are connected to the aperture mass estimators.

B1 Set partitions and Bell polynomials

We begin by defining a partition π of a set $n = \{1, 2,..., n\}$ as a collection of mutually exclusive subsets (blocks) of n whose union equals n. In our case, all these partitions can be mapped on to an associated partition λ being defined as the number of elements of each block in π. Each element λ can be represented as $(n_1, n_2,..., n_m)$ or as $(1^{m_1},\ 2^{m_2},\ \cdots,\ n^{m_n})$ where for the former expression the n_i denote the length of the ith block while for the latter case the m_i represent the number of occurrences of a block of length i in π. If π is a partition of n having m blocks this implies that $\sum_i m_i = m$ and $\sum_i i\, m_i = n$. We will now show that the following proposition holds: Proposition:

For the set n and a partition λ of length m given as $(1^{m_1},\ 2^{m_2},\ \cdots,\ \ell^{m_\ell})$, there are $\frac{n!}{\prod_{i=1}^{\ell} m_i!(i!)^{m_i}}$ partitions π of n having the same $\lambda(\pi)$.

Proof**:**

As a first step, we just look at the number of ways the m subsets can be chosen from $\boldsymbol{n}$. This can easily be worked out when noting that for the first subset there are $\binom{n}{n_1}$ choices, for the following $\binom{n-n_1}{n_2}$, etc. Following through all of the subsets, we then have

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots n_{m-2}}{n_{m-1}}\binom{n_m}{n_m} = \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-\cdots n_{m-2})!}{n_{m-1}!n_m!}\frac{n_m!}{n_m!} = \frac{n!}{n_1!n_2!\cdots n_m!} \tag{B1}$$

possibilities. Shifting this expression to the representation of λ given above we see that many of them give the identical partition π; to get rid of those ones we need to divide by the number of ways all the equal size blocks themselves can be permuted with each other. Applying those conditions we have

$$\frac{1}{\text{Norm.}} \times \frac{n!}{n_1!n_2!\cdots n_m!} = \frac{1}{m_1!m_2!\cdots m_\ell!} \times \frac{n!}{(1!)^{m_1}(2!)^{m_2}\cdots(\ell!)^{m_\ell}} = \frac{n!}{\prod_{i=1}^{\ell} m_i!(i!)^{m_i}} \tag{B2}$$

possibilities remaining, which is exactly the proposed expression.

With this result in hand we are now in position to understand the form of the partial Bell polynomial $B_{n,\,m}$ being defined as

$$B_{n,m}(x_1\cdots, x_{n-m+1}) = \sum_{(m_1,\cdots,m_{n-m+1})\in P_{n,m}} \frac{n!}{m_1!\cdots m_{n-m+1}!}\left(\frac{x_1}{1!}\right)^{m_1}\cdots\left(\frac{x_{n-m+1}}{(n-m+1)!}\right)^{m_{n-m+1}}, \tag{B3}$$

where

$$P_{n,m} \equiv \left\{(m_1,\cdots m_{n-m+1}) \in \mathbb{N}_0^{n-m+1} \,\middle|\, \sum_{i=1}^{n-m+1} m_i = m,\quad \sum_{i=1}^{n-m+1} i\, m_i = n\right\}.$$

Comparing the prefactors and the index set[13] with our discussion above we see that the partial Bell polynomials simply sum over all the partitions λ of $\mathbf{n}$ having a fixed m, i.e. they list the number of ways a set consisting the n objects can be partitioned into m blocks. For example, looking at $B_{4,\,2}$ the allowed index combinations are $\{(0, 2, 0), (1, 0, 1)\}$ such that equation (B3) evaluates to $B_{4,2} = 4x_1x_3 + 3x_2^2$. We note in the passing that these expressions generate the same prefactors that arise in the halo model, i.e. we can relate the structure of $B_{4,\,2}$ to the two-halo term of the halo model trispectrum.

[13]The upper limit is given by the partition having the largest possible block size, namely $(1^{m\,-\,1}, 2^0, \cdots, (n-(m-1))^1)$

Finally, we define the complete Bell polynomial B_n which list all possible partitions of n objects:

$$B_n(x_1 \cdots, x_{n-m+1}) = \sum_{m=1}^{n} B_{n,m}(x_1 \cdots, x_{n-m+1}) = \sum_{m=1}^{n} \sum_{\pi \in P_{n,m}} \prod_{i=1}^{n-m(\lambda(\pi))+1} x_i^{m_i(\lambda(\pi))}\,, \tag{B4}$$

where the first equality states the formal definition and the second one rewrites it into an explicit sum over all the partitions of the set $\boldsymbol{n}$.

B2 Sums over unequal indices and Bell polynomials

Let us look at the simple expression $\sum_{i=1}^{N}\sum_{j\neq i}^{N} x_i x_j$. A naive implementation of this double sum would imply a quadratic complexity of the corresponding program. A much faster way resulting in linear complexity can be achieved when noting that $(\sum_{i_1=1}^{N} x_{i_1})(\sum_{i_2=1}^{N} x_{i_2}) = \sum_{i_1=1}^{N}\sum_{i_2\neq i_1}^{N} x_{i_1}x_{i_2} + \sum_{i_1=1}^{N} x_1^2$. We can easily generalize this pattern by treating the number of indices as the set n from the previous subsection. Then all the different partitions λ of this set correspond to different ways these indices can be set equal with one another; the corresponding prefactors can be obtained via the Bell polynomial. To clarify this statement, we write down as an example the expression for $n = 4$:

$$\begin{aligned}\left(\sum_i x_i\right)^4 &= \sum_{i_1\neq i_2\neq i_3\neq i_4} x_{i_1}x_{i_2}x_{i_3}x_{i_4} + \left(\sum_{i_1\neq i_3\neq i_4} x_{i_1}^2 x_{i_3}x_{i_4} + 5\text{perm.}\right) + \left(\sum_{i_1\neq i_2} x_{i_1}^3 x_{i_2} + 3\text{perm.}\right) + \left(\sum_{i_1\neq i_3} x_{i_1}^2 x_{i_3}^2 + 2\text{perm.}\right) + \sum_{i_1} x_{i_1}^4 \\ &\sim \left(1^4,\ 2^0,\ 3^0,\ 4^0\right) + 6\times\left(1^2,\ 2^1,\ 3^0,\ 4^0\right) + 4\times\left(1^1,\ 2^0,\ 3^1,\ 4^0\right) + 3\times\left(1^0,\ 2^2,\ 3^0,\ 4^0\right) + \left(1^0,\ 2^0,\ 3^0,\ 4^1\right).\end{aligned} \tag{B5}$$

From here, we see that we can express a sum over n unequal indices in terms of two power sums and a set of related sums over at most $n-1$ unequal indices. Repeating the same argument on the latter sums one eventually arrives at an expression only involving power sums. Carrying out aforementioned calculations along the lines of Appendix A for our example this yields

$$\sum_{i_1\neq i_2\neq i_3\neq i_4} x_{i_1}x_{i_2}x_{i_3}x_{i_4} = \left(\sum_i x_i\right)^4 - 6\left(\sum_i x_i\right)^2\left(\sum_i x_i^2\right) + 8\left(\sum_i x_i\right)\left(\sum_i x_i^3\right) + 3\left(\sum_i x_i^2\right)^2 - 6\left(\sum_i x_i^4\right)\,. \tag{B6}$$

Comparing the latter two expressions, we note that their index partitions are the same, but that they differ in some signs and prefactors; namely there is a negative sign for an odd partition length m and an additional multiplicative factor of $(i-1)!$ for each block of length i. Looking at the structure of equation (B6), i.e. the fact that all of its summands correspond to a partition of an integer set and that furthermore it constitutes of n different building blocks we might be tempted to cast it in terms of Bell polynomials with the identifying the x_ℓ from equation (B4) with the power sums $c_\ell \sum_i x_i^\ell \;:\; c_\ell \in \mathbb{R}$. In the next paragraphs, we formalize these observations and from there determine the c_ℓ.

The first difference can be motivated most easily by choosing a graphical representation in which we draw each index as a single point. Then the prefactors in equation (B5) are given by the number of ways one can group together different points such that they constitute the corresponding partition whereas for equation (B6) it additionally matters in which order these points have been set equal with each other, which in mathematical terms is described by how many closed cycles one can draw between them. The induced correction of $(\ell-1)!$ for a block of length ℓ can be absorbed in the Bell polynomial by setting $c_\ell = (\ell-1)!$.

The second observation can be generalized inductively. Looking at our example of $n = 4$ we see that the sign for each partition λ is given by $\mathrm{sgn}(\lambda) = \prod_{i=1}^{n}(-1)^{m_i(\lambda)((i+1)\bmod 2)}$, that is each block of even length contributes a negative sign. Performing the induction step, we have

$$\sum_{i_1\neq\cdots i_n\neq i_{n+1}} x_{i_1}\cdots x_{i_{n+1}} = \left(\sum_{i_{n+1}} x_{i_{n+1}}\right)\left(\sum_{i_1\neq\cdots i_3\neq i_n} x_{i_1}\cdots x_{i_n}\right) - \left[\left(\sum_{i_1\neq\cdots i_3\neq i_n} x_{i_1}^2 x_{i_2}\cdots x_{i_n}\right) + (n-1)\text{ perm.}\right]\,. \tag{B7}$$

Looking at the modification of the partitions, for the first term we have $m_1 \to m_1 + 1$ for all λ such that we would not have expected any sign flips. For the second term, we need to update the block in which the identical index sits, assuming it had length k we have $m_k \to m_k - 1$ and $m_{k+1} \to m_{k+1} + 1$. In case of an even k reducing its occurrence by one induces an additional sign flip whereas for odd ks we get a sign flip for the increase of m_{k+1}. Putting things together we conclude that we could predict the correct signs by examining the partition structures. Therefore, setting $c_\ell = (-1)^{(\ell+1)\bmod 2}(\ell-1)!$ in equation (B4) will reproduce generalizations of *equation* (*B*6). We can brush this in a nicer shape by setting $c_\ell = -(\ell-1)!$ and furthermore multiplying B_n by $(-1)^n$; this modification effectively just multiplies each term of the previous result by an even power of negative one.

With these two modifications in hand we can finally write down the main result of this subsection, namely the way on how to transform a sum over unequal indices into a sum over products of power sums:

$$\sum_{i_1\neq\cdots\neq i_n} x_{i_1}\cdots x_{i_n} = (-1)^n\, B_n\left(-0!\sum_i x_i,\ -1!\sum_i x_i^2, \cdots,\ -(n-1)!\sum_i x_i^n\right) \tag{B8}$$

B3 Application to the aperture mass estimator

Looking at the form of equation (B8), the expression for the direct estimator of the aperture statistics with equal aperture radii equation (24) immediately follows when identifying the arguments in the nominator and denominator with the power sums $M_{\mathrm{s},m}$ and S_m and cancelling the overall sign.

For the case of unequal aperture radii, we still need to do a bit more work. Looking back to our previous example equation (B5), having unequal aperture radii induces different values of the Q filters such that the x_i cannot be taken to be the same variable anymore. Hence we have to replace the prefactors in equation (B5) by a sum over all the possible ways the different radii can be partitioned. The second set of prefactors that arises when going to equation (B6) still applies in the case of unequal radii as it effectively corresponds to swapping two aperture radii in the corresponding multivariate power sum equation (31). Thus, it seems appropriate to formulate the solution via summing over partitions, such that we can rewrite equation (29) as

$$\widehat{M^n_{\rm ap}}(\vartheta_1, ..., \vartheta_n) = \frac{\sum_{m=1}^{n}\sum_{\pi\in P_{n,m}}(-1)^m\prod_{i=1}^{m}(n_i-1)!\; M^{(n_i)}_{\rm s,(\jmath_1(\pi_i),\cdots,\jmath_n(\pi_i))}}{\sum_{m=1}^{n}\sum_{\pi\in P_{n,m}}(-1)^m\prod_{i=1}^{m}(n_i-1)!\; S^{(n_i)}_{(\jmath_1(\pi_i),\cdots,\jmath_n(\pi_i))}}\,. \tag{B9}$$

We note that from this formulation one can build an efficient way of computing equation (32) within the subset $\mathcal{U}$ of the datacube $[R_1, \cdots R_m]^n$ ($m \geq n$) in which neither of the indices are equal: This is due to the fact that the number of power sums in which $1 \leq i \leq n$ radii are selected is simply given by $\binom{m}{i}$ and therefore the n-dimensional hypercube of aperture radii can be constructed from a set consisting of just $\sum_{i=1}^{n}\binom{m}{i}$ power sums. After allocating those power sums for all the galaxies within an aperture we can then enumerate through the relevant aperture radii multiplets, select the relevant subsets of the power sums, and then again apply the transformation equation (32) to transform to the multiscale aperture mass moments, or equivalently to their corresponding connected parts. With the help of this procedure, we were able to conduct the full analysis displayed in Fig. 9 on the SLICS ensemble (a total of around 2.5 billion galaxies) within just 6000 CPU hours.

B4 Expressions of the accelerated estimator for low orders (unequal radii)

In order to save space, we only write down the expressions for the nominator or equation (32), the denominator will have an identical structure. As expected, the number of sums in the nth-order estimator equals the nth Bell number.

$$\widehat{M^1_{\rm ap}}(\vartheta_1) = \frac{1}{\rm norm}\times M^{(1)}_{\rm s,(1)}\,, \tag{B10}$$

$$\widehat{M^2_{\rm ap}}(\vartheta_1,\vartheta_2) = \frac{1}{\rm norm}\times\left\{M^{(1)}_{\rm s,(1,0)}M^{(1)}_{\rm s,(0,1)} - M^{(2)}_{\rm s,(1,1)}\right\}\,, \tag{B11}$$

$$\widehat{M^3_{\rm ap}}(\vartheta_1,\vartheta_2,\vartheta_3) = \frac{1}{\rm norm}\times\left\{M^{(1)}_{\rm s,(1,0,0)}M^{(1)}_{\rm s,(0,1,0)}M^{(1)}_{\rm s,(0,0,1)} - \left[M^{(2)}_{\rm s,(1,1,0)}\,M^{(1)}_{\rm s,(0,0,1)} + 2\text{ perm.}\right] + 2\,M^{(3)}_{\rm s,(1,1,1)}\right\}\,, \tag{B12}$$

$$\begin{aligned}\widehat{M^4_{\rm ap}}(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4) = \frac{1}{\rm norm}\times\Big\{&M^{(1)}_{\rm s,(1,0,0,0)}M^{(1)}_{\rm s,(0,1,0,0)}M^{(1)}_{\rm s,(0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,1)} - \left[M^{(2)}_{\rm s,(1,1,0,0)}M^{(1)}_{\rm s,(0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,1)} + 5\text{ perm.}\right]\\ &+\left[M^{(2)}_{\rm s,(1,1,0,0)}M^{(2)}_{\rm s,(0,0,1,1)} + 2\text{ perm.}\right] + 2\left[M^{(3)}_{\rm s,(1,1,1,0)}M^{(1)}_{\rm s,(0,0,0,1)} + 3\text{ perm.}\right] - 6\,M^{(4)}_{\rm s,(1,1,1,1)}\Big\}\,,\end{aligned} \tag{B13}$$

$$\begin{aligned}\widehat{M^5_{\rm ap}}(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4,\vartheta_5) = \frac{1}{\rm norm}\times\Big\{&M^{(1)}_{\rm s,(1,0,0,0,0)}M^{(1)}_{\rm s,(0,1,0,0,0)}M^{(1)}_{\rm s,(0,0,1,0,0)}M^{(1)}_{\rm s,(0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1)}\\ &-\left[M^{(2)}_{\rm s,(1,1,0,0,0)}M^{(1)}_{\rm s,(0,0,1,0,0)}M^{(1)}_{\rm s,(0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1)} + 9\text{ perm.}\right]\\ &+\left[M^{(2)}_{\rm s,(1,1,0,0,0)}M^{(2)}_{\rm s,(0,0,1,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1)} + 14\text{ perm.}\right]\\ &+2\left[M^{(3)}_{\rm s,(1,1,1,0,0)}M^{(1)}_{\rm s,(0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1)} + 9\text{ perm.}\right] - 2\left[M^{(3)}_{\rm s,(1,1,1,0,0)}M^{(2)}_{\rm s,(0,0,0,1,1)} + 9\text{ perm.}\right]\\ &+6\left[M^{(4)}_{\rm s,(1,1,1,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1)} + 4\text{ perm.}\right] + 24\,M^{(5)}_{\rm s,(1,1,1,1,1)}\Big\}\,,\end{aligned} \tag{B14}$$

$$\begin{aligned}\widehat{M^6_{\rm ap}}(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4,\vartheta_5,\vartheta_6) = \frac{1}{\rm norm}\times\Big\{&M^{(1)}_{\rm s,(1,0,0,0,0,0)}M^{(1)}_{\rm s,(0,1,0,0,0,0)}M^{(1)}_{\rm s,(0,0,1,0,0,0)}M^{(1)}_{\rm s,(0,0,0,1,0,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,0,1)}\\ &-\left[M^{(2)}_{\rm s,(1,1,0,0,0,0)}M^{(1)}_{\rm s,(0,0,1,0,0,0)}M^{(1)}_{\rm s,(0,0,0,1,0,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)} + 14\text{ perm.}\right]\\ &+\left[M^{(2)}_{\rm s,(1,1,0,0,0,0)}M^{(2)}_{\rm s,(0,0,1,1,0,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,0,1)} + 44\text{ perm.}\right]\\ &-\left[M^{(2)}_{\rm s,(1,1,0,0,0,0)}M^{(2)}_{\rm s,(0,0,1,1,0,0)}M^{(2)}_{\rm s,(0,0,0,0,1,1)} + 14\text{ perm.}\right]\\ &+2\left[M^{(3)}_{\rm s,(1,1,1,0,0,0)}M^{(1)}_{\rm s,(0,0,0,1,0,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)} + 19\text{ perm.}\right]\\ &-2\left[M^{(3)}_{\rm s,(1,1,1,0,0,0)}M^{(2)}_{\rm s,(0,0,0,1,1,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)} + 59\text{ perm.}\right]\\ &+4\left[M^{(3)}_{\rm s,(1,1,1,0,0,0)}M^{(3)}_{\rm s,(0,0,0,1,1,1)} + 9\text{ perm.}\right]\\ &-6\left[M^{(4)}_{\rm s,(1,1,1,1,0,0)}M^{(1)}_{\rm s,(0,0,0,0,1,0)}M^{(1)}_{\rm s,(0,0,0,0,0,1)} + 14\text{ perm.}\right]\\ &+6\left[M^{(4)}_{\rm s,(1,1,1,1,0,0)}M^{(2)}_{\rm s,(0,0,0,0,1,1)} + 14\text{ perm.}\right] + 24\left[M^{(5)}_{\rm s,(1,1,1,1,1,0)}M^{(1)}_{\rm s,(0,0,0,0,0,1)} + 5\text{ perm.}\right]\\ &-120\,M^{(6)}_{\rm s,(1,1,1,1,1,1)}\Big\}\,.\end{aligned} \tag{B15}$$

APPENDIX C: VARIANCE OF THE DIRECT ESTIMATOR

C1 Motivation of the shape and multiplicity factor

We recall the definition of the $\widehat{M_{\rm ap}^n}$ variance:

$$\sigma^2\left[\widehat{M_{\rm ap}^n}\right] = \mathbb{E}\left[\left(\widehat{M_{\rm ap}^n}\right)^2\right] - \left\langle \mathcal{M}_{\rm ap}^n\right\rangle^2 = \frac{\left(\pi\vartheta^2\right)^{2n}}{\left(\sum_{\neq} w_{j_1}\cdots w_{j_n}\right)^2}\cdot\mathbb{E}\left[\sum_{\neq} w_{i_1}\cdots w_{i_n}x_{i_1}\cdots x_{i_n}\cdot\sum_{\neq} w_{j_1}\cdots w_{j_n}x_{j_1}\cdots x_{j_n}\right] - \left\langle \mathcal{M}_{\rm ap}^n\right\rangle^2 , \quad \text{(C1)}$$

where we defined $x_i \equiv Q_i e_{t,\,i}$ for notational simplicity. We proceed along the standard lines by decomposing the expectation value in an averaging step A over the intrinsic ellipticity distribution, another one P over the galaxy positions, and finally one over the cosmological ensemble. Let us start by applying the ellipticity averaging procedure for which $A(e_i, e_j) \equiv \frac{\sigma_\epsilon^2}{2}\delta_{i,j}^K + \gamma_i\gamma_j\left(1-\delta_{i,j}^K\right)$. Noting that each summation sign in (C1) runs over an index set where all the indices are unequal, we see that only indices between the two sums can be contracted to yield the shape noise expression. We can represent the index structure graphically as $|\,i_1\cdots i_n\,|\,j_1\cdots j_n|$ and define a contraction as a line between two indices of the i and j set. The prefactor of the term in the A-averaging is then given by the number of possible contractions.

As an example, let us compute the prefactor when applying two contractions in the variance of the third-order statistics. For the first contraction, there are nine possibilities, while for each second one there are only for indices remaining, giving four further possibilities. As the contractions are interchangeable we need to divide the result by two to yield a prefactor of 18. A graphical representation of this explanation would look as follows:

$$\left|\overbrace{i_1\;\overbrace{i_2\;i_3}\;|\;j_1\;\overbrace{j_2\;j_3}}\right| = \frac{9}{2}\times\left|i_2\;\overbrace{i_3\;|\;j_2}\;j_3\right| = \frac{9\cdot 4}{2!} = 18\,.$$

This scheme allows us to easily generalize our example to performing ℓ contractions on the nth-order statistics, giving a prefactor of $C_2(n,\ell) \equiv \frac{n^2(n-1)^2\cdots(n-\ell-1)^2}{\ell!}$.

For the position averaging, we can repeat the same argument, as $P(Q_i\gamma_t Q_j\gamma_j) \sim M_{\rm s,2}\delta_{i,j}^K + M_{\rm ap}^2\left(1-\delta_{i,j}^K\right)$. If we already have performed ℓ contractions for the A-averaging, there are only $(n-\ell)$ free indices left in each block – hence there will be $C_2(n-\ell, p)$ possibilities to perform p additional contractions in the P-averaging.

Next, we compute the expectation value for a given index set in which we have performed ℓ contractions in the A-averaging and p contractions in the P-averaging:

$$\begin{aligned}
&\Bigg\langle P\left(\sum_{\neq} w_{i_1}^2 Q_{i_1}^2\cdots w_{i_\ell}^2 Q_{i_\ell}^2\; w_{i_{\ell+1}}^2 Q_{i_{\ell+1}}^2\gamma_{t,i_{\ell+1}}^2\cdots w_{i_{\ell+p}}^2 Q_{i_{\ell+p}}^2\gamma_{t,i_{\ell+p}}^2\right.\\
&\qquad\left.w_{i_{\ell+p+1}}Q_{i_{\ell+p+1}}\gamma_{t,i_{\ell+p+1}}\cdots w_{i_n}Q_{i_n}\gamma_{t,i_n}\; w_{j_{\ell+p+1}}Q_{j_{\ell+p+1}}\gamma_{t,j_{\ell+p+1}}\cdots w_{j_n}Q_{i_n}\gamma_{t,j_n}\right)\Bigg\rangle\\
&\quad\equiv\Bigg\langle\prod_{i=1}^N\int_{\rm Ap.}\frac{{\rm d}^2\theta_i}{\pi\vartheta^2}\sum_{\neq} w_{i_1}^2 Q_{i_1}^2\cdots w_{i_\ell}^2 Q_{i_\ell}^2\; w_{i_{\ell+1}}^2 Q_{i_{\ell+1}}^2\gamma_{t,i_{\ell+1}}^2\cdots w_{i_{\ell+p}}^2 Q_{i_{\ell+p}}^2\gamma_{t,i_{\ell+p}}^2\\
&\qquad w_{i_{\ell+p+1}}Q_{i_{\ell+p+1}}\gamma_{t,i_{\ell+p+1}}\cdots w_{i_n}Q_{i_n}\gamma_{t,i_n}\; w_{j_{\ell+p+1}}Q_{j_{\ell+p+1}}\gamma_{t,j_{\ell+p+1}}\cdots w_{j_n}Q_{i_n}\gamma_{t,j_n}\Bigg\rangle .
\end{aligned}$$

$$\begin{aligned}
&= \sum_{/} = w_{i_1}^2\cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}}\cdots w_{i_n}w_{j_{\ell+p+1}}\cdots w_{j_n}\\
&\quad\left(\prod_{i\in\{i_1,\cdots,i_\ell\}}\int_{\rm Ap.}\frac{{\rm d}^2\theta_i}{\pi\vartheta^2}Q_i^2\right)\left\langle\left(\prod_{i\in\{i_{\ell+1},\cdots,i_{\ell+p}\}}\int_{\rm Ap.}\frac{{\rm d}^2\theta_i}{\pi\vartheta^2}Q_i^2\gamma_{t,i}^2\right)\left(\prod_{i\in\{i_{\ell+p+1},\cdots,j_n\}}\int_{\rm Ap.}\frac{{\rm d}^2\theta_i}{\pi\vartheta^2}Q_i\gamma_{t,i}\right)\right\rangle\left(\int_{\rm Ap.}\frac{{\rm d}^2\theta_i}{\pi\vartheta^2}\right)^{N-2(\ell+p)}\\
&= \sum_{/} = w_{i_1}^2\cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}}\cdots w_{i_n}w_{j_{\ell+p+1}}\cdots w_{j_n}\;\times\;\prod_{i=1}^{\ell}\left(\int_{\rm Ap.}\frac{{\rm d}^2\theta_i}{\pi\vartheta^2}Q_i^2\right)\left\langle\prod_{j=1}^{p}\left(\int_{\rm Ap.}\frac{{\rm d}^2\theta_j}{\pi\vartheta^2}Q_j^2\gamma_j^2\right)^{2(n-\ell-p)}\prod_{k=1}^{2(n-\ell-p)}\left(\int_{\rm Ap.}\frac{{\rm d}^2\theta_k}{\pi\vartheta^2}Q_k\gamma_k\right)\right\rangle\\
&\quad\equiv\frac{\sum_{/} w_{i_1}^2\cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}}\cdots w_{i_n}w_{j_{\ell+p+1}}\cdots w_{j_n}}{(\pi\vartheta)^{2n}}\;\times\;M_{g,2}^{\ell}\left\langle\mathcal{M}_{s,2}^p\mathcal{M}_{\rm ap}^{2(n-\ell-p)}\right\rangle .
\end{aligned}$$

Note that in this derivation the order of the contracted indices does not matter as they all end up to be integration variables. If we now combine this result together with the multiplicity factors we can write a closed form expresson for (C1):

$$\begin{aligned}\sigma^2\left[\widehat{M_{\rm ap}^n}\right] &= \sum_{\ell=0}^{n} C_2(n,\ell)\left(\frac{\sigma_\epsilon^2}{2}\right)^\ell M_{g,2}^\ell \sum_{p=0}^{n-\ell} \frac{\sum_{\neq} w_{i_1}^2\cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}}\cdots w_{i_n} w_{j_{\ell+p+1}}\cdots w_{j_n}}{\left(\sum_{\neq} w_{i_1}\cdots w_{i_n}\right)^2} C_2(n-\ell,p)\left\langle \mathcal{M}_{s,2}^p \mathcal{M}_{\rm ap}^{2(n-\ell-p)}\right\rangle - \left\langle \mathcal{M}_{\rm ap}^n\right\rangle^2 \\ &\approx \sum_{\ell=0}^{n} \frac{\sum_{\neq} w_{i_1}^2\cdots w_{i_\ell}^2 w_{i_{\ell+1}}\cdots w_{i_n} w_{j_{\ell+1}}\cdots w_{j_n}}{\left(\sum_{\neq} w_{i_1}\cdots w_{i_n}\right)^2}\,\ell!\binom{n}{\ell}\binom{n}{\ell}\left(\frac{\sigma_\epsilon^2}{2}\right)^\ell M_{g,2}^\ell \left\langle \mathcal{M}_{\rm ap}^{2(n-\ell)}\right\rangle - \left\langle \mathcal{M}_{\rm ap}^n\right\rangle^2 \\ &\approx n!\frac{\sum_{\neq} w_{i_1}^2\cdots w_{i_n}^2}{\left(\sum_{\neq} w_{i_1}\cdots w_{i_n}\right)^2}\left(\frac{\sigma_\epsilon^2}{2}\right)^n M_{g,2}^n\,. \end{aligned} \tag{C2}$$

The first line is equivalent to (35) when combining the multiplicity factors and adjusting the indices. The second line makes the approximation that each of the $\mathcal{M}_{\rm s,2}$ are negligible (which is true for large N); for the final line, we only keep the shot noise contribution.

C2 Modifications for unequal aperture radii

In case of multiple apertures, the structure of the variance is basically unchanged, the only thing we need to adjust is to use the multivariate version of the power sums and to replace the multiplicity factor with a sum over the actual multivariate expressions such that their radii correspond to the structure of the contracted indices. If we then take the shot noise dominated case we end up with:

$$\sigma^2_{\rm shot}\left[\widehat{M_{\rm ap}^n}(R_1,\cdots,R_n)\right] = \frac{\sum_{/=}' w_{i_1}^2\cdots w_{i_n}^2}{\left(\sum_{/=}' w_{i_1}\cdots w_{i_n}\right)^2}\left(\frac{\sigma_\epsilon^2}{2}\right)^n \sum_{\beta_1\neq\cdots\neq\beta_n}\prod_{i=1}^{n} G_2\left(\frac{\max\left(\{R_i,R_{\beta_i}\}\right)}{\min\left(\{R_i,R_{\beta_i}\}\right)}\right)\,, \tag{C3}$$

where we define G_2 as the multiple radii generalization of $\mathcal{M}_{\rm g,2}$:

$$G_2(\beta) \equiv \pi R^2\int {\rm d}^2\boldsymbol{\theta}\; Q_R(\theta)Q_{\beta R}(\theta) = \frac{72}{\beta^2}\left[\frac{1}{24}-\frac{1}{8\beta^2}+\frac{1}{10\beta^4}\right] \quad (\beta\geq 1)$$

where the second equality denotes the corresponding equation for the polynomial filter. Note that for the corresponding inverse shot noise weighting scheme only the sum over the weights matters, as the remainder of the above expression is constant and can be factored out.

C3 Explicit expressions for low orders

Here, we collect the lowest order explicit expressions for (35). The second-order expression was first derived in Schneider (1998). Note that our prefactors differ from the ones defined in (Munshi & Coles 2003).

$$\sigma^2\left[\widehat{\mathcal{M}_{\rm ap}}\right] = \frac{1}{\left(\sum_{/=} w_{i_1}\right)^2}\left\{\sum_{/=} w_{i_1}w_{j_1}\left\langle\mathcal{M}_{\rm ap}^2\right\rangle + \sum_{/=} w_{i_1}^2\left\langle\mathcal{M}_{s,2}\right\rangle + 1\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)\sum_{/=} w_{i_1}^2\right\} - \left\langle\mathcal{M}_{\rm ap}\right\rangle^2$$

$$\begin{aligned}\sigma^2\left[\widehat{\mathcal{M}_{\rm ap}^2}\right] = \frac{1}{\left(\sum_{/=} w_{i_1}w_{i_2}\right)^2}\Bigg\{&\sum_{/=} w_{i_1}w_{j_1}w_{i_2}w_{j_2}\left\langle\mathcal{M}_{\rm ap}^4\right\rangle + 4\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^2\right\rangle + 2\sum_{/=} w_{i_1}^2 w_{i_2}^2\left\langle\mathcal{M}_{s,2}^2\right\rangle \\ &+4\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)\left[\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}\left\langle\mathcal{M}_{\rm ap}^2\right\rangle + \sum_{/=} w_{i_1}^2 w_{i_2}^2\left\langle\mathcal{M}_{s,2}\right\rangle\right] + 2\mathcal{M}_{g,2}^2\left(\frac{\sigma_\epsilon^2}{2}\right)^2\sum_{/=} w_{i_1}^2 w_{i_2}^2\Bigg\} - \left\langle\mathcal{M}_{\rm ap}^2\right\rangle^2\end{aligned}$$

$$\begin{aligned}\sigma^2\left[\widehat{\mathcal{M}_{\rm ap}^3}\right] = \frac{1}{\left(\sum_{/=} w_{i_1}w_{i_2}w_{i_3}\right)^2}\Bigg\{&\sum_{/=} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}_{\rm ap}^6\right\rangle + 9\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^4\right\rangle \\ &+18\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^2\right\rangle + 6\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2\left\langle\mathcal{M}_{s,2}^3\right\rangle + 9\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)\Bigg[\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}_{\rm ap}^4\right\rangle \\ &+4\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^2\right\rangle + 2\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2\left\langle\mathcal{M}_{s,2}^2\right\rangle\Bigg] + 18\mathcal{M}_{g,2}^2\left(\frac{\sigma_\epsilon^2}{2}\right)^2\Bigg[\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}\left\langle\mathcal{M}_{\rm ap}^2\right\rangle \\ &+\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2\left\langle\mathcal{M}_{s,2}\right\rangle\Bigg] + 6\mathcal{M}_{g,2}^3\left(\frac{\sigma_\epsilon^2}{2}\right)^3\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2\Bigg\} - \left\langle\mathcal{M}_{\rm ap}^3\right\rangle^2\end{aligned}$$

$$\sigma^2\left[\widehat{\mathcal{M}_{\rm ap}^4}\right] = \frac{1}{\left(\sum_{/=} w_{i_1}w_{i_2}w_{i_3}w_{i_4}\right)^2}\left\{\sum_{/=} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle\mathcal{M}_{\rm ap}^8\right\rangle + 16\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^6\right\rangle\right.$$
$$+72\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^4\right\rangle + 96\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle\mathcal{M}_{s,2}^3\mathcal{M}_{\rm ap}^2\right\rangle + 24\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle\mathcal{M}_{s,2}^4\right\rangle$$
$$+16\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)\left[\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle\mathcal{M}_{\rm ap}^6\right\rangle + 9\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^4\right\rangle\right.$$
$$\left.+18\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^2\right\rangle + 6\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle\mathcal{M}_{s,2}^3\right\rangle\right] + 72\mathcal{M}_{g,2}^2\left(\frac{\sigma_\epsilon^2}{2}\right)^2$$
$$\left[\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle\mathcal{M}_{\rm ap}^4\right\rangle + 4\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^2\right\rangle + 2\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle\mathcal{M}_{s,2}^2\right\rangle\right]$$
$$\left.+96\mathcal{M}_{g,2}^3\left(\frac{\sigma_\epsilon^2}{2}\right)^3\left[\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle\mathcal{M}_{\rm ap}^2\right\rangle + \sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle\mathcal{M}_{s,2}\right\rangle\right] + 24\mathcal{M}_{g,2}^4\left(\frac{\sigma_\epsilon^2}{2}\right)^4\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\right\} - \left\langle\mathcal{M}_{\rm ap}^4\right\rangle^2$$

$$\sigma^2\left[\widehat{\mathcal{M}_{\rm ap}^5}\right] = \frac{1}{\left(\sum_{/=} w_{i_1}w_{i_2}w_{i_3}w_{i_4}w_{i_5}\right)^2}\left\{\sum_{/=} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{\rm ap}^{10}\right\rangle\right.$$
$$+25\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^8\right\rangle + 200\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^6\right\rangle$$
$$+600\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}^3\mathcal{M}_{\rm ap}^4\right\rangle + 600\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}^4\mathcal{M}_{\rm ap}^2\right\rangle$$
$$+120\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2\left\langle\mathcal{M}_{s,2}^5\right\rangle + 25\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)\left[\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{\rm ap}^8\right\rangle\right.$$
$$+16\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^6\right\rangle + 72\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^4\right\rangle$$
$$+96\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}^3\mathcal{M}_{\rm ap}^2\right\rangle + 24\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2\left\langle\mathcal{M}_{s,2}^4\right\rangle\right] + 200\mathcal{M}_{g,2}^2\left(\frac{\sigma_\epsilon^2}{2}\right)^2$$
$$\left[\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{\rm ap}^6\right\rangle + 9\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^4\right\rangle\right.$$
$$+18\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^2\right\rangle + 6\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2\left\langle\mathcal{M}_{s,2}^3\right\rangle\right] + 600\mathcal{M}_{g,2}^3\left(\frac{\sigma_\epsilon^2}{2}\right)^3$$
$$\left[\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle\mathcal{M}_{\rm ap}^4\right\rangle + 4\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^2\right\rangle + 2\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2\left\langle\mathcal{M}_{s,2}^2\right\rangle\right]$$
$$+600\mathcal{M}_{g,2}^4\left(\frac{\sigma_\epsilon^2}{2}\right)^4\left[\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}\left\langle\mathcal{M}_{\rm ap}^2\right\rangle + \sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2\left\langle\mathcal{M}_{s,2}\right\rangle\right]$$
$$\left.+120\mathcal{M}_{g,2}^5\left(\frac{\sigma_\epsilon^2}{2}\right)^5\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2\right\} - \left\langle\mathcal{M}_{\rm ap}^5\right\rangle^2$$

$$\sigma^2\left[\widehat{\mathcal{M}_{\rm ap}^6}\right] = \frac{1}{\left(\sum_{/=} w_{i_1}w_{i_2}w_{i_3}w_{i_4}w_{i_5}w_{i_6}\right)^2}\left\{\sum_{/=} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}w_{i_6}w_{j_6}\left\langle\mathcal{M}_{\rm ap}^{12}\right\rangle\right.$$
$$+36\sum_{/=} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}w_{i_6}w_{j_6}\left\langle\mathcal{M}_{s,2}\mathcal{M}_{\rm ap}^{10}\right\rangle + 450\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}w_{i_6}w_{j_6}\left\langle\mathcal{M}_{s,2}^2\mathcal{M}_{\rm ap}^8\right\rangle$$
$$+2400\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}w_{i_5}w_{j_5}w_{i_6}w_{j_6}\left\langle\mathcal{M}_{s,2}^3\mathcal{M}_{\rm ap}^6\right\rangle + 5400\sum_{/=} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}w_{i_6}w_{j_6}\left\langle\mathcal{M}_{s,2}^4\mathcal{M}_{\rm ap}^4\right\rangle$$

$$+4320\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^5 \mathcal{M}_{\rm ap}^2 \right\rangle + 720\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^6 \right\rangle + 36\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)$$

$$\left[\sum_{/} = w_{i_1}^2 w_{i_2} w_{j_2} w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\rm ap}^{10} \right\rangle + 25\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\rm ap}^8 \right\rangle\right.$$

$$+200\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\rm ap}^6 \right\rangle + 600\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^3 \mathcal{M}_{\rm ap}^4 \right\rangle$$

$$\left.+600\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^4 \mathcal{M}_{\rm ap}^2 \right\rangle + 120\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^5 \right\rangle\right] + 450\mathcal{M}_{g,2}^2\left(\frac{\sigma_\epsilon^2}{2}\right)^2$$

$$\left[\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\rm ap}^8 \right\rangle + 16\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\rm ap}^6 \right\rangle\right.$$

$$+72\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\rm ap}^4 \right\rangle + 96\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^3 \mathcal{M}_{\rm ap}^2 \right\rangle$$

$$\left.+24\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^4 \right\rangle\right] + 2400\mathcal{M}_{g,2}^3\left(\frac{\sigma_\epsilon^2}{2}\right)^3 \left[\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\rm ap}^6 \right\rangle\right.$$

$$+9\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\rm ap}^4 \right\rangle + 18\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\rm ap}^2 \right\rangle$$

$$\left.+6\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^3 \right\rangle\right] + 5400\mathcal{M}_{g,2}^4\left(\frac{\sigma_\epsilon^2}{2}\right)^4 \left[\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\rm ap}^4 \right\rangle\right.$$

$$\left.+4\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\rm ap}^2 \right\rangle + 2\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^2 \right\rangle\right] + 4320\mathcal{M}_{g,2}^5\left(\frac{\sigma_\epsilon^2}{2}\right)^5$$

$$\left[\sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\rm ap}^2 \right\rangle + \sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2} \right\rangle\right] + 720\mathcal{M}_{g,2}^6\left(\frac{\sigma_\epsilon^2}{2}\right)^6 \sum_{/} = w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \Bigg\}$$

$$-\left\langle \mathcal{M}_{\rm ap}^6 \right\rangle^2$$

This paper has been typeset from a TEX/LATEX file prepared by the author.